STIC-Biotech/ChemLib

5982/

From:	Hunt, Jennifer
Sent:	Wednesday, February 06, 2002 7:28 PM
To:	STIC-Biotech/ChemLib
Subject:	Seq Search for 09/480,977

Please search and interference search SEQ ID NO:4 of 09/480,977.

Thanks.

Jennifer Hunt
Patent Examiner, Art Unit 1642
CM1-8D06 (mailbox 8E12)
(703)308-7548

Edward Hart
Technical Info. Specialist
STIC / Biotech
CM1 12C14 Tel: 305-9203

09/8/? ,647
09/877,665 michael park —5/1/09
101 132, 577
101 132, 200

us-09-480-977-4.rai

OM protein - protein search, using sw model

Copyright (c) 1993 - 2000 Compugen Ltd.

GenCore version 4.5

Run on: February 7, 2002, 12:56:46 ; Search time 12.54 Seconds
                                              (without alignments)
                                              84.342 Million cell updates/sec

Title: 277
Perfect score: 277
Sequence: 1 HFKPCRDKDLAYCLNDGECF............SHKHCRCKEGYQGVRCDQFL 47

Scoring table:
Scoring matrix: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_AA:*
                 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
                 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
                 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
                 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
                 5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
                 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	277	100.0	47	3	US-08-899-437-4	Sequence 4, Appl
2	277	100.0	47	3	US-08-899-437-4	Sequence 4, Appl
3	277	100.0	47	3	US-09-126-121-4	Sequence 8, Appl
4	277	100.0	47	3	US-08-899-437-8	Sequence 8, Appl
5	277	100.0	47	3	US-09-126-121-8	Sequence 4, Appl
6	277	100.0	360	4	US-08-899-437-7	Sequence 7, Appl
7	277	100.0	362	4	US-09-126-121-7	Sequence 7, Appl
8	277	100.0	402	3	US-08-899-437-3	Sequence 3, Appl
9	277	100.0	697	4	US-09-126-121-3	Sequence 3, Appl
10	277	100.0	1696	4	US-08-899-437-23	Sequence 23, Appl
11	277	100.0	1696	4	US-09-126-121-23	Sequence 23, Appl
12	277	100.0	713	4	US-08-899-437-2	Sequence 2, Appl
13	277	100.0	720	4	US-09-126-121-2	Sequence 2, Appl
14	277	100.0	620	3	US-08-899-437-6	Sequence 6, Appl
15	116.5	42.1	52	1	US-09-126-121-6	Sequence 6, Appl
16	116.5	42.1	52	1	US-08-417-640A-1	Sequence 1, Appl
17	116.5	42.1	52	1	US-08-760-815-1	Sequence 1, Appl
18	116.5	42.1	52	2	US-09-238-182-1	Sequence 1, Appl
19	113.5	41.0	49	3	US-08-899-437-14	Sequence 14, Appl
20	113.5	41.0	49	4	US-09-126-121-14	Sequence 14, Appl
21	113.5	41.0	50	4	US-08-753-007A-12	Sequence 12, Appl
22	113.5	41.0	50	4	US-08-398-496-12	Sequence 12, Appl
23	113.5	41.0	52	1	US-08-417-640A-3	Sequence 3, Appl
24	113.5	41.0	52	1	US-08-760-815-3	Sequence 3, Appl
25	113.5	41.0	52	2	US-08-761-038-3	Sequence 3, Appl
26	113.5	41.0	54	2	US-08-179-481-111	Sequence 111, Appl
27	113.5	41.0	63	3	US-08-341-018-62	Sequence 62, Appl
28	113.5	41.0	63	4	US-08-470-335-221	Sequence 221, Appl
29	113.5	41.0	66	4	US-08-470-339-221	Sequence 221, App
30	113.5	41.0	66	1	US-07-847-743B-10	Sequence 10, Appl
31	113.5	41.0	66	1	US-08-456-201-10	Sequence 10, Appl
32	113.5	41.0	66	4	US-08-341-018-68	Sequence 68, Appl
33	113.5	41.0	66	2	US-09-020-880-2	Sequence 2, Appl1
34	113.5	41.0	66	5	PCT-US92-04295A-10	Sequence 10, Appl
35	113.5	41.0	83	4	US-08-341-018-70	Sequence 70, Appl
36	113.5	41.0	83	4	US-08-470-335-225	Sequence 225, App
37	113.5	41.0	83	4	US-08-470-339-225	Sequence 225, App
38	113.5	41.0	88	3	US-08-341-018-68	Sequence 68, Appl
39	113.5	41.0	88	4	US-08-470-335-224	Sequence 224, App
40	113.5	41.0	95	1	US-08-470-339-224	Sequence 224, Appl
41	113.5	41.0	95	1	US-07-847-743B-14	Sequence 14, Appl
42	113.5	41.0	95	1	US-08-456-201-14	Sequence 14, Appl
43	113.5	41.0	95	2	US-08-330-161-14	Sequence 14, Appl
44	113.5	41.0	95	2	US-08-456-241-14	Sequence 14, Appl
45	113.5	41.0	95	2	US-08-440-401-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-899-437-4
Sequence 4, Application US/08899437
Patent No. 6121415
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; TITLE OF INVENTION: Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,437
; FILING DATE: 24-Jul-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 47 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: NRG3 EGF-like domain/amino acid seq.
; LOCATION: 1-47
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
US-08-899-437-4

Qy   1 HFKPCRDKDLAYCLNDGECFVIEILTGSHKHCRCKEGYQGVRCDQFL 47

Query Match   100.0%; Score 277; DB 3; Length 47;
Best Local Similarity 100.0%; Pred. No. 1.2e-26;
Matches   47; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Db
US-08-899-437-8
1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
   |||||||||||||||||||||||||||||||||||||||||||||||

RESULT 2
US-08-899-437-8
; Sequence 8, Application US/0889943?
; Patent No. 6121415
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; TITLE OF INVENTION: Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,437
; FILING DATE: 24-Jul-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 47 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: NRG3 EGF-like domain/amino acid seq.
; LOCATION: 1-47
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
US-08-899-437-8

Query Match   100.0%;   Score 277; DB 3;   Length 47;
Best Local Similarity   100.0%;   Pred. No. 1.2e-26;
Matches   47; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
               |||||||||||||||||||||||||||||||||||||||||||||||
Db   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47

RESULT 3
US-09-126-121-4
; Sequence 4, Application US/09126121
; Patent No. 6252051
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; TITLE OF INVENTION: Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA

ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/126,121
FILING DATE: 30-Jul-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P1084R1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
FEATURE:
NAME/KEY: NRG3 EGF-like domain/amino acid seq.
LOCATION: 1-47
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-09-126-121-4

Query Match   100.0%;   Score 277; DB 4;   Length 47;
Best Local Similarity   100.0%;   Pred. No. 1.2e-26;
Matches   47; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
               |||||||||||||||||||||||||||||||||||||||||||||||
Db   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47

RESULT 4
US-09-126-121-8
; Sequence 8, Application US/09126121
; Patent No. 6252051
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; TITLE OF INVENTION: Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/126,121
; FILING DATE: 30-Jul-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:

```
;   LENGTH:  47 amino acids
;   TYPE:  Amino Acid
;   TOPOLOGY:  Linear
;   FEATURE:
;     NAME/KEY:  NRG3 EGF-like domain/amino acid seq.
;     LOCATION:  1-47
;     IDENTIFICATION METHOD:
;     OTHER INFORMATION:
US-09-126-121-8

Qy      1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
          |||||||||||||||||||||||||||||||||||||||||||||||
Db      1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47

Query Match             100.0%;  Score 277;  DB 4;  Length 47;
Best Local Similarity   100.0%;  Pred. No. 1.2e-26;
Matches    47;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

RESULT 5
US-08-899-437-7
; Sequence 7, Application US/0889437
; Patent No. 6121415
; GENERAL INFORMATION:
;   APPLICANT:  Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
;   TITLE OF INVENTION:  ErbB Receptor-Specific Neuregulin Related
;   TITLE OF INVENTION:  Ligands and Uses Therefor
;   NUMBER OF SEQUENCES:  23
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Genentech, Inc.
;     STREET:  1 DNA Way
;     CITY:  South San Francisco
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  94080
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  WinPatin (Genentech)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/899,437
;     FILING DATE:  24-Jul-1997
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Conley, Deirdre L.
;     REGISTRATION NUMBER:  36,487
;     REFERENCE/DOCKET NUMBER:  P1084R1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  650/225-2066
;     TELEFAX:  650/952-9881
; INFORMATION FOR SEQ ID NO:  7:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  360 amino acids
;     TYPE:  Amino Acid
;     TOPOLOGY:  Linear
;   FEATURE:
;     NAME/KEY:  hNRG3 extracellular domain/Amino AcidSeq
;     LOCATION:  1-360
;     IDENTIFICATION METHOD:
;     OTHER INFORMATION:
US-08-899-437-7

Qy      1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    286 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 332

Query Match             100.0%;  Score 277;  DB 3;  Length 360;
Best Local Similarity   100.0%;  Pred. No. 9.6e-26;
Matches    47;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
RESULT 6
US-09-126-121-7
; Sequence 7, Application US/09126121
; Patent No. 6252051
; GENERAL INFORMATION:
;   APPLICANT:  Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
;   TITLE OF INVENTION:  ErbB Receptor-Specific Neuregulin Related
;   TITLE OF INVENTION:  Ligands and Uses Therefor
;   NUMBER OF SEQUENCES:  23
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Genentech, Inc.
;     STREET:  1 DNA Way
;     CITY:  South San Francisco
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  94080
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  WinPatin (Genentech)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/126,121
;     FILING DATE:  30-Jul-1998
;     CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Conley, Deirdre L.
;     REGISTRATION NUMBER:  36,487
;     REFERENCE/DOCKET NUMBER:  P1084R1D1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  650/225-2066
;     TELEFAX:  650/952-9881
; INFORMATION FOR SEQ ID NO:  7:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  360 amino acids
;     TYPE:  Amino Acid
;     TOPOLOGY:  Linear
;   FEATURE:
;     NAME/KEY:  hNRG3 extracellular domain/Amino AcidSeq
;     LOCATION:  1-360
;     IDENTIFICATION METHOD:
;     OTHER INFORMATION:
US-09-126-121-7

Qy      1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    286 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 332

Query Match             100.0%;  Score 277;  DB 4;  Length 360;
Best Local Similarity   100.0%;  Pred. No. 9.6e-26;
Matches    47;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

RESULT 7
US-08-899-437-3
; Sequence 3, Application US/0889437
; Patent No. 6121415
; GENERAL INFORMATION:
;   APPLICANT:  Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
;   TITLE OF INVENTION:  ErbB Receptor-Specific Neuregulin Related
;   TITLE OF INVENTION:  Ligands and Uses Therefor
;   NUMBER OF SEQUENCES:  23
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Genentech, Inc.
;     STREET:  1 DNA Way
;     CITY:  South San Francisco
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  94080
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
```

```
             COMPUTER: IBM PC compatible
             OPERATING SYSTEM: PC-DOS/MS-DOS
             SOFTWARE: WinPatin (Genentech)
           CURRENT APPLICATION DATA:
             APPLICATION NUMBER: US/08/899,437
             FILING DATE: 24-Jul-1997
             CLASSIFICATION: 435
           ATTORNEY/AGENT INFORMATION:
             NAME: Conley, Deirdre L.
             REGISTRATION NUMBER: 36,487
             REFERENCE/DOCKET NUMBER: P1084R1
           TELECOMMUNICATION INFORMATION:
             TELEPHONE: 650/225-2066
             TELEFAX: 650/952-9881
         INFORMATION FOR SEQ ID NO: 3:
           SEQUENCE CHARACTERISTICS:
             LENGTH: 362 amino acids
             TYPE: Amino Acid
             TOPOLOGY: Linear
           FEATURE:
             NAME/KEY: mNRG3 extracellular domainAmino acid seq
             LOCATION: 1-362
             IDENTIFICATION METHOD:
             OTHER INFORMATION:
 US-08-899-437-3
```

```
Qy    1  HFKPCRDKDLAYCLNDGECFVIETLLTGSHKHCRCKEGYQGVRCDQFL 47
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  288  HFKPCRDKDLAYCLNDGECFVIETLLTGSHKHCRCKEGYQGVRCDQFL 334
```

```
Query Match            100.0%;  Score 277;  DB 3;  Length 362;
Best Local Similarity  100.0%;  Pred. No. 9.6e-26;
Matches  47;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
RESULT   8
US-09-126-121-3
Sequence 3, Application US/09126121
Patent No. 6252051
GENERAL INFORMATION:
  APPLICANT: Godowski, Paul J. Mark, Melanie Rose, Zhang, Dong Xiao
  TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
  TITLE OF INVENTION: Ligands and Uses Therefor
  NUMBER OF SEQUENCES: 23
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Genentech, Inc.
    STREET: 1 DNA Way
    CITY: South San Francisco
    STATE: California
    COUNTRY: USA
    ZIP: 94080
  COMPUTER READABLE FORM:
    MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: WinPatin (Genentech)
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/126,121
    FILING DATE: 30-Jul-1998
    CLASSIFICATION:
  ATTORNEY/AGENT INFORMATION:
    NAME: Conley, Deirdre L.
    REGISTRATION NUMBER: 36,487
    REFERENCE/DOCKET NUMBER: P1084R1D1
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 650/225-2066
    TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 3:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 362 amino acids
    TYPE: Amino Acid
    TOPOLOGY: Linear
```

```
           FEATURE:
             NAME/KEY: mNRG3 extracellular domainAmino acid seq
             LOCATION: 1-362
             IDENTIFICATION METHOD:
             OTHER INFORMATION:
 US-09-126-121-3
```

```
Qy    1  HFKPCRDKDLAYCLNDGECFVIETLLTGSHKHCRCKEGYQGVRCDQFL 47
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  288  HFKPCRDKDLAYCLNDGECFVIETLLTGSHKHCRCKEGYQGVRCDQFL 334
```

```
Query Match            100.0%;  Score 277;  DB 4;  Length 362;
Best Local Similarity  100.0%;  Pred. No. 9.6e-26;
Matches  47;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
RESULT   9
US-08-899-437-23
Sequence 23, Application US/08899437
Patent No. 6121415
GENERAL INFORMATION:
  APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
  TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
  TITLE OF INVENTION: Ligands and Uses Therefor
  NUMBER OF SEQUENCES: 23
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Genentech, Inc.
    STREET: 1 DNA Way
    CITY: South San Francisco
    STATE: California
    COUNTRY: USA
    ZIP: 94080
  COMPUTER READABLE FORM:
    MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: WinPatin (Genentech)
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/899,437
    FILING DATE: 24-Jul-1997
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Conley, Deirdre L.
    REGISTRATION NUMBER: 36,487
    REFERENCE/DOCKET NUMBER: P1084R1
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 650/225-2066
    TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 23:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 696 amino acids
    TYPE: Amino Acid
    TOPOLOGY: Linear
  FEATURE:
    NAME/KEY: Human NRG3B2
    LOCATION: 1-696
    IDENTIFICATION METHOD:
    OTHER INFORMATION:
US-08-899-437-23
```

```
Qy    1  HFKPCRDKDLAYCLNDGECFVIETLLTGSHKHCRCKEGYQGVRCDQFL 47
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  286  HFKPCRDKDLAYCLNDGECFVIETLLTGSHKHCRCKEGYQGVRCDQFL 332
```

```
Query Match            100.0%;  Score 277;  DB 3;  Length 696;
Best Local Similarity  100.0%;  Pred. No. 1.9e-25;
Matches  47;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
RESULT   10
US-09-126-121-23
```

Sequence 23, Application US/09126121
Patent No. 6252051
GENERAL INFORMATION:
;   APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
;   TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
;   TITLE OF INVENTION: Ligands and Uses Therefor
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genentech, Inc.
;   STREET: 1 DNA Way
;   CITY: South San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94080
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: WinPatin (Genentech)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/126,121
;   FILING DATE: 30-Jul-1998
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Conley, Deirdre L.
;   REGISTRATION NUMBER: 36,487
;   REFERENCE/DOCKET NUMBER: P1084R1D1
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 650/225-2066
;   TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 23:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 696 amino acids
;   TYPE: Amino Acid
;   TOPOLOGY: Linear
;   FEATURE:
;   NAME/KEY: Human NRG3B2
;   LOCATION: 1-696
;   IDENTIFICATION METHOD:
;   OTHER INFORMATION:
US-09-126-121-23

Query Match   100.0%; Score 277; DB 4;   Length 696;
Best Local Similarity   100.0%; Pred. No. 1.9e-25;
Matches   47; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
        |||||||||||||||||||||||||||||||||||||||||||||||
Db 286 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 332

RESULT 11
US-08-899-437-2
Sequence 2, Application US/0889437
Patent No. 612141S
GENERAL INFORMATION:
;   APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
;   TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
;   TITLE OF INVENTION: Ligands and Uses Therefor
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genentech, Inc.
;   STREET: 1 DNA Way
;   CITY: South San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94080
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: WinPatin (Genentech)

CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/899,437
;   FILING DATE: 24-Jul-1997
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Conley, Deirdre L.
;   REGISTRATION NUMBER: 36,487
;   REFERENCE/DOCKET NUMBER: P1084R1
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 650/225-2066
;   TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 713 amino acids
;   TYPE: Amino Acid
;   TOPOLOGY: Linear
;   FEATURE:
;   NAME/KEY: Mouse NRG3 (mNRG3)/amino acid seq.
;   LOCATION: 1-713
;   IDENTIFICATION METHOD:
;   OTHER INFORMATION:
US-08-899-437-2

Query Match   100.0%; Score 277; DB 3;   Length 713;
Best Local Similarity   100.0%; Pred. No. 1.9e-25;
Matches   47; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47.
        |||||||||||||||||||||||||||||||||||||||||||||||
Db 288 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 334

RESULT 12
US-09-126-121-2
Sequence 2, Application US/09126121
Patent No. 6252051
GENERAL INFORMATION:
;   APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
;   TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
;   TITLE OF INVENTION: Ligands and Uses Therefor
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genentech, Inc.
;   STREET: 1 DNA Way
;   CITY: South San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94080
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: WinPatin (Genentech)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/126,121
;   FILING DATE: 30-Jul-1998
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Conley, Deirdre L.
;   REGISTRATION NUMBER: 36,487
;   REFERENCE/DOCKET NUMBER: P1084R1D1
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 650/225-2066
;   TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 713 amino acids
;   TYPE: Amino Acid
;   TOPOLOGY: Linear
;   FEATURE:
;   NAME/KEY: Mouse NRG3 (mNRG3)/amino acid seq.
;   LOCATION: 1-713

IDENTIFICATION METHOD:
OTHER INFORMATION:
US-09-126-121-2

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
       |||||||||||||||||||||||||||||||||||||||||||||||
Db 288 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 334

Query Match   100.0%;   Score 277;   DB 4;   Length 713;
Best Local Similarity   100.0%;   Pred. No. 1.9e-25;
Matches   47;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 13
US-08-899-437-6
Sequence 6, Application US/08899437
Patent No. 6121415
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
TITLE OF INVENTION: Ligands and Uses Therefor
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,437
FILING DATE: 24-Jul-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P1084R1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 720 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
FEATURE:
NAME/KEY: hNRG3B1 amino acid sequence
LOCATION: 1-720
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-08-899-437-6

Query Match   100.0%;   Score 277;   DB 3;   Length 720;
Best Local Similarity   100.0%;   Pred. No. 1.9e-25;
Matches   47;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
       |||||||||||||||||||||||||||||||||||||||||||||||
Db 286 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 332

RESULT 14
US-09-126-121-6
Sequence 6, Application US/09126121
Patent No. 6252051
GENERAL INFORMATION:

APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
TITLE OF INVENTION: Ligands and Uses Therefor
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/126,121
FILING DATE: 30-Jul-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P1084R1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 720 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
FEATURE:
NAME/KEY: hNRG3B1 amino acid sequence
LOCATION: 1-720
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-09-126-121-6

Query Match   100.0%;   Score 277;   DB 4;   Length 720;
Best Local Similarity   100.0%;   Pred. No. 1.9e-25;
Matches   47;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
       |||||||||||||||||||||||||||||||||||||||||||||||
Db 286 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 332

RESULT 15
US-08-417-640A-1
Sequence 1, Application US/0417640A
Patent No. 5670342
GENERAL INFORMATION:
APPLICANT: Amgen Inc.
APPLICANT: Carnahan, Josette F.
APPLICANT: Hara, Shinichi
APPLICANT: Lu, Hsieng S.
APPLICANT: Mayer, John P.
APPLICANT: Yoshinaga, Steven K.
TITLE OF INVENTION: NDF Peptides
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

```
CURRENT APPLICATION DATA:
    APPLICATION NUMBER:  US/08/417,640A
    FILING DATE:
    CLASSIFICATION:  436
ATTORNEY/AGENT INFORMATION:
    NAME:  Mazza, Richard J.
    REFERENCE/DOCKET NUMBER:  A-310
INFORMATION FOR SEQ ID NO:  1:
SEQUENCE CHARACTERISTICS:
    LENGTH:  52 amino acids
    TYPE:  amino acid
    STRANDEDNESS:  single
    TOPOLOGY:  linear
    MOLECULE TYPE:  protein
US-08-417-640A-1
```

```
Query Match          42.1%;  Score 116.5;  DB 1;  Length 52;
Best Local Similarity   33.3%;  Pred. No. 1.7e-07;
Matches    16;  Conservative   15;  Mismatches   16;  Indels    1;  Gaps    1;

Qy    1  HFKFCRDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL 47
            |   :|:| |:|||::|:| |:  :| |:| : | || :  ::
Db    2  HLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCKCQPGFTGARCQNYV 49
```

```
Search completed: February 7, 2002, 13:00:10
Job time: 204 sec
```

```
                    GenCore version 4.5
              Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        February 7, 2002, 12:54:46 ; Search time 23.88 Seconds
                                            (without alignments)
                                            145.789 Million cell updates/sec

Title:         US-09-480-977-4
Perfect score: 277
Sequence:      1 HFKPCRDKDLAYCLNDGECF..........SHKHCRCKEGYQGVRCDQFL 47
Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5
Searched:      522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters:    522463
Post-processing: Minimum Match 08
                 Maximum Match 100%
Minimum DB seq length: 0
Maximum DB seq length: 200000000
                 Listing first 45 summaries

Database :     A_Geneseq_1101:*
         1:    /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
         2:    /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
         3:    /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
         4:    /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
         5:    /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
         6:    /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
         7:    /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
         8:    /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
         9:    /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
        10:    /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
        11:    /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
        12:    /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
        13:    /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
        14:    /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
        15:    /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
        16:    /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
        17:    /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
        18:    /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
        19:    /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
        20:    /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
        21:    /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
        22:    /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	277	100.0	47	20	AAW97622	Human neuregulin r
2	277	100.0	157	20	AAX05451	Human heregulin-li
3	277	100.0	360	20	AAW97621	Human neuregulin r
4	277	100.0	362	20	AAW97620	Mouse neuregulin r
5	277	100.0	696	20	AAW97619	Human neuregulin r
6	277	100.0	713	20	AAW97619	Mouse neuregulin r
7	277	100.0	720	20	AAW97618	Human neuregulin r
8	277	100.0	720	20	AAW05452	Human neuregulin r
9	277	100.0	720	20	AAX05452	Human neuregulin-l1
10	116.5	42.1	52	20	AAW05182	Human neuregulin r
10	116.5	42.1	52	17	AAB12602	Neu differentiatio
11	116.5	42.1	52	21	AAX69983	NDF/heregulin prot
12	113.5	41.0	52	17	AAW05184	Neu differentiatio
13	113.5	41.0	63	15	AAR46918	EGFL2. AAR46918;
14	113.5	41.0	63	16	AAR56659	EGFL2. AAR56659;
15	113.5	41.0	63	16	AAR67250	Human epidermal 11
16	113.5	41.0	63	17	AAW03363	EGF12. Not specif
17	113.5	41.0	63	17	AAW09076	Epidermal growth f
18	113.5	41.0	66	21	AAB36702	EGF-like domain of
19	113.5	41.0	63	17	AAR46922	Epidermal growth f
20	113.5	41.0	83	15	AAR55663	EGFL6. AAR55663;
21	113.5	41.0	83	17	AAR55663	EGFL6. AAR55663;
22	113.5	41.0	83	16	AAR67254	Human epidermal 11
23	113.5	41.0	83	17	AAW03367	EGFL6. Not specif
24	113.5	41.0	83	17	AAR96080	Epidermal growth f
25	113.5	41.0	83	17	AAR87465	Epidermal growth f
26	113.5	41.0	88	15	AAR56921	EGFL5. AAR56921;
27	113.5	41.0	88	15	AAR55662	EGFL5. AAR55662;
28	113.5	41.0	88	16	AAR67253	Human epidermal 11
29	113.5	41.0	88	17	AAW03366	EGFL5. Not specif
30	113.5	41.0	88	17	AAR96079	Epidermal growth f
31	113.5	41.0	88	17	AAR87464	Epidermal growth f
32	113.5	41.0	125	16	AAR66564	Human NDF-alpha3
33	113.5	41.0	263	13	AAR28537	GGF2BPP2.CDS prote
34	113.5	41.0	263	15	AAR46896	GGF2BPP2. Bos tau
35	113.5	41.0	263	15	AAR55689	GGF2BPP2. Bos tau
36	113.5	41.0	263	16	AAR67217	Putative bovine gl
37	113.5	41.0	263	17	AAW09360	Bovine neuregulin
38	113.5	41.0	280	15	AAR46915	Epidermal growth f
39	113.5	41.0	280	16	AAR57244	Bovine glial cell
40	113.5	41.0	280	16	AAR57244	Bovine glial cell
41	113.5	41.0	280	17	AAW09369	Human neuregulin G
42	113.5	41.0	280	17	AAW98762	Glial growth facto
43	113.5	41.0	280	17	AAR87455	BPP2 glial growth
44	113.5	41.0	280	20	AAY26577	Amino acid sequenc
45	113.5	41.0	375	18	AAW06683	Heregulin-alpha fu

ALIGNMENTS

```
RESULT    1
AAW97622
AAW97622 standard; Protein; 47 AA.

ID   AAW97622 standard; Protein; 47 AA.
XX
AC   AAW97622;
XX
DT   10-MAY-1999  (first entry)
XX
DE   Human neuregulin related ligand NRG3 EGF-like domain.
XX
KW   Neuregulin related ligand; NRG3; hNRG3B1; human; Erbb4 receptor;
KW   signal transduction; nervous system disorder; neurodegeneration;
KW   neuropathy; therapy; diagnosis; epidermal growth factor; EGF;
KW   immunoadhesin.
OS   Homo sapiens.
XX
PN   WO9902681-A1.
XX
PD   21-JAN-1999.
XX
PF   30-JUN-1998;  98WO-US13411.
XX
PR   24-JUL-1997;  97US-0899437.
PR   09-JUL-1997;  97US-0052019.
XX
PA   (GETH ) GENENTECH INC.
PI   Godowski PJ, Mark MR, Zhang D;
XX
DR   WPI; 1999-120882/10.
XX
PT   New isolated neuregulin related ligand-3 - used to develop products
PT   for treating nervous system disorders, e.g. stroke, ischaemia,
```

PT   Infection, malignancy, Alzheimer's disease or Down's syndrome
XX
PS   Claim 30; Page 64; 101pp; English.
XX
CC   This is the epidermal growth factor (EGF)-like domain of human
CC   neuregulin related ligand NRG3 (see also AAW97618), a novel member of
CC   the EGF-like family of protein ligands that binds to the ErbB4
CC   receptor and activates ErbB4 receptor tyrosine phosphorylation.
CC   The EGF-1 like domain of NRG3 is distinct from the EGF-like domains
CC   of NRG1 and NRG2. The invention provides human and murine
CC   polypeptides (see also AAW97617) that have at least 75% homology to
CC   the NRG3 EGF-like domain, as well as expression vectors, host cells
CC   and methods for the recombinant production of novel NRG3s. The
CC   NRG3 polypeptides and polynucleotides and can be used to enhance
CC   the survival, proliferation or differentiation of cells having the
CC   ErbB4 receptor in vivo and in vitro. They can be used to prevent
CC   or treat damage to a nerve or damage to other NRG3-expressing or
CC   NRG3-responsive cells, e.g. brain, heart, or kidney cells. In
CC   particular, they can be used to treat diseases which involve neural
CC   cell growth such as demyelination, or damage or loss of glial cells
CC   (e.g. multiple sclerosis). They can be used to treat patients whose
CC   nervous system has been damaged by e.g. trauma, surgery, stroke,
CC   ischemia, infection, metabolic disease, nutritional deficiency,
CC   malignancy, or toxic agents. NRG3 can also be used to treat
CC   skeletal or smooth muscle wasting. The products can also be used
CC   for detection, diagnosis, for the production of transgenic or
CC   knockout animals or for drug screening. A claimed immunoadhesin
CC   comprises the human NRG3 EGF-like domain fused to an immunoglobulin
CC   sequence.
XX
SQ   Sequence   47 AA;

```
Query Match             100.0%;  Score 277;  DB 20;  Length 47;
Best Local Similarity   100.0%;  Pred. No. 1.7e-21;
Matches    47;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
              |||||||||||||||||||||||||||||||||||||||||||||||
Db          1 hfkpcrdkdlayclndgecfvietltgshkhcrckegyqgvrcdqfl 47
```

RESULT   2
AAY05451
ID   AAY05451 standard; Protein; 157 AA.
XX
AC   AAY05451;
XX
DT   06-JUL-1999 (first entry)
XX
DE   Human heregulin-like factor sequence.
XX
KW   Human heregulin-like factor; HLF; cell growth regulator; diagnosis;
KW   neural system disorder; cancer.
XX
OS   Homo sapiens.
XX
PN   WO9857989-A1.
XX
PD   23-DEC-1998.
XX

PF   16-JUN-1998; 98WO-US12403.
XX
PR   17-JUN-1997; 97US-0049942.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (GEOU) UNIV GEORGETOWN.
XX
PI   Hijazi MM, King CR, Ruben SM, Young P;
XX
DR   WPI; 1999-095327/08.
DR   N-PSDB; AAX36423.
XX
PT   New isolated heregulin-like factor - used to develop products for
PT   the diagnosis and treatment of disorders involving regulation of
PT   cell growth, particularly cancers
XX
PS   Claim 17; Page 86-87; 118pp; English.
XX
CC   This sequence is the human heregulin-like factor (HLF) of the
CC   invention. The HLF is involved in the regulation of cell growth.
CC   Detection of different levels of expression of the HLF gene can be used
CC   for the diagnosis of disorders, e.g. in the neural system. In
CC   particular, detection of different levels of HLF gene expression in cells
CC   or body fluid of an individual can be used for diagnosing cancer. The
CC   products can also be used in the treatment of disorders involving
CC   abnormal levels of HLF activity.
XX
SQ   Sequence   157 AA;

```
Query Match             100.0%;  Score 277;  DB 20;  Length 157;
Best Local Similarity   100.0%;  Pred. No. 5.1e-21;
Matches    47;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         31 hfkpcrdkdlayclndgecfvietltgshkhcrckegyqgvrcdqfl 77
```

RESULT   3
AAW97621
ID   AAW97621 standard; Protein; 360 AA.
XX
AC   AAW97621;
XX
DT   10-MAY-1999 (first entry)
XX
DE   Human neuregulin related ligand NRG3 extracellular domain.
XX
KW   Neuregulin related ligand; NRG3; hNRG3B1; human; ErbB4 receptor;
KW   signal transduction; nervous system disorder; neurodegeneration;
KW   neuropathy; therapy; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO9902681-A1.
XX
PD   21-JAN-1999.
XX
PF   30-JUN-1998; 98WO-US13411.
XX
PR   24-JUL-1997; 97US-0899437.
PR   09-JUL-1997; 97US-0052019.
XX
PA   (GETH) GENENTECH INC.
XX
PI   Godowski PJ, Mark MR, Zhang D;
XX
DR   WPI; 1999-120882/10.
XX
PT   New isolated neuregulin related ligand-3 - used to develop products
PT   for treating nervous system disorders, e.g. stroke, ischaemia,
PT   Infection, malignancy, Alzheimer's disease or Down's syndrome

XX
PS Claim 5(a); Page 69-70; 101pp; English.
XX
CC This is the extracellular domain (ECD, aa1-360 of human neuregulin
CC related ligand NRG3 (see also AAW97618), a novel member of the
CC epidermal growth factor (EGF)-like family of protein ligands. NRG3
CC binds to the ErbB4 receptor, but not to the ErbB2 or ErbB3 receptor,
CC and activates ErbB4 receptor tyrosine phosphorylation. The invention
CC provides human and murine polypeptides (see also AAW97617) that have
CC at least 75% homology to the NRG3 ECD, as well as expression vectors,
CC host cells and methods for the recombinant production of novel
CC NRG3s. The NRG3 polypeptides and polynucleotides and can be used to
CC enhance the survival, proliferation or differentiation of cells
CC having the ErbB4 receptor in vivo and in vitro. They can be used to
CC prevent or treat damage to a nerve or damage to other NRG3-expressing
CC or NRG3-responsive cells, e.g. brain, heart, or kidney cells. In
CC particular, they can be used to treat diseases which involve neural
CC cell growth such as demyelination, or damage or loss of glial cells
CC (e.g. multiple sclerosis). They can be used to treat patients whose
CC nervous system has been damaged by e.g. trauma, surgery, stroke,
CC ischaemia, infection, metabolic disease, nutritional deficiency,
CC malignancy, or toxic agents. NRG3 can also be used to treat
CC motor neuron disorders such as amyotrophic lateral sclerosis (Lou
CC Gehrig's disease), Bell's palsy, conditions involving spinal
CC muscular atrophy or paralysis, neurodegenerative disorders such as
CC Alzheimer's disease, Parkinson's disease, epilepsy, multiple
CC sclerosis, Huntington's chorea, Down's syndrome, nerve deafness,
CC and Meniere's disease. They can also be used to treat neuropathies
CC associated with systemic disease including post-polio syndrome,
CC hereditary neuropathies including Charcot-Marie-Tooth disease,
CC Refsum's disease, abetalipoproteinemia, Tangier disease, Krabbe's
CC disease, metachromatic leukodystrophy, Fabry's disease and
CC Dejerine-Sottas syndrome, to treat disease of skeletal muscle of
CC smooth muscle, such as muscular dystrophy or diseases caused by
CC skeletal or smooth muscle wasting. The products can also be used
CC for detection, diagnosis, for the production of transgenic or
CC knockout animals or for drug screening.
XX
SQ Sequence 360 AA;

Query Match 100.0%; Score 277; DB 20; Length 360;
Best Local Similarity 100.0%;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
  |||||||||||||||||||||||||||||||||||||||||||||||
Db 286 hfkpcrdkdlayclndgecfvietltgshkhcrckegyqgvrcdqfl 332

RESULT 4
ID AAW97620 standard; Protein; 362 AA.
XX
AC AAW97620;
XX
DT 10-MAY-1999 (first entry)
XX
DE Mouse neuregulin related ligand NRG3 extracellular domain.
XX
OS Mus sp.
XX
KW Neuregulin related ligand; NRG3; mouse; ErbB4 receptor;
KW signal transduction; nervous system disorder; neurodegeneration;
KW neuropathy; therapy; diagnosis.
XX
PN WO902681-A1.
XX
PD 21-JAN-1999.
XX
PF 30-JUN-1998; 98WO-US13411.
XX
PR 24-JUL-1997; 97US-0899437.

PR 09-JUL-1997; 97US-0052019.
XX
PA (GETH) GENENTECH INC.
XX
PI Godowski PJ, Mark MR, Zhang D;
XX
DR WPI: 1999-120882/10.
XX
PT New isolated neuregulin related ligand-3 - used to develop products
PT for treating nervous system disorders, e.g. stroke, ischaemia,
PT infection, malignancy, Alzheimer's disease or Down's syndrome
XX
PS Claim 5(a); Page 62-63; 101pp; English.
XX
CC This is the extracellular domain (ECD, aa1-362) of murine neuregulin
CC related ligand NRG3 (see also AAW97617), a novel member of the
CC epidermal growth factor (EGF)-like family of protein ligands. NRG3
CC binds to the ErbB4 receptor, but not to the ErbB2 or ErbB3 receptor,
CC activates ErbB4 receptor tyrosine phosphorylation. The invention
CC provides human and murine polypeptides (see also AAW97618) that have
CC at least 75% homology to the NRG3 ECD, as well as expression vectors,
CC host cells and methods for the recombinant production of novel
CC NRG3s. The NRG3 polypeptides and polynucleotides and can be used to
CC enhance the survival, proliferation or differentiation of cells
CC having the ErbB4 receptor in vivo and in vitro. They can be used to
CC prevent or treat damage to a nerve or damage to other NRG3-expressing
CC or NRG3-responsive cells, e.g. brain, heart, or kidney cells. In
CC particular, they can be used to treat diseases which involve neural
CC cell growth such as demyelination, or damage or loss of glial cells
CC (e.g. multiple sclerosis). They can be used to treat patients whose
CC nervous system has been damaged by e.g. trauma, surgery, stroke,
CC ischaemia, infection, metabolic disease, nutritional deficiency,
CC malignancy, or toxic agents. NRG3 can also be used to treat
CC motor neuron disorders such as amyotrophic lateral sclerosis (Lou
CC Gehrig's disease), Bell's palsy, conditions involving spinal
CC muscular atrophy or paralysis, neurodegenerative disorders such as
CC Alzheimer's disease, Parkinson's disease, epilepsy, multiple
CC sclerosis, Huntington's chorea, Down's syndrome, nerve deafness,
CC and Meniere's disease. They can also be used to treat neuropathies
CC associated with systemic disease including post-polio syndrome,
CC hereditary neuropathies including Charcot-Marie-Tooth disease,
CC Refsum's disease, abetalipoproteinemia, Tangier disease, Krabbe's
CC disease, metachromatic leukodystrophy, Fabry's disease and
CC Dejerine-Sottas syndrome, to treat disease of skeletal muscle of
CC smooth muscle, such as muscular dystrophy or diseases caused by
CC skeletal or smooth muscle wasting. The products can also be used
CC for detection, diagnosis, for the production of transgenic or
CC knockout animals or for drug screening.
XX
SQ Sequence 362 AA;

Query Match 100.0%; Score 277; DB 20; Length 362;
Best Local Similarity 100.0%;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
  |||||||||||||||||||||||||||||||||||||||||||||||
Db 288 hfkpcrdkdlayclndgecfvietltgshkhcrckegyqgvrcdqfl 334

RESULT 5
ID AAW97619 standard; Protein; 696 AA.
XX
AC AAW97619;
XX
DT 10-MAY-1999 (first entry)
XX
DE Human neuregulin related ligand NRG3 (splice variant).
XX
KW Neuregulin related ligand; NRG3; hNRG3B1; human; ErbB4 receptor;
KW signal transduction; nervous system disorder; neurodegeneration;

Left column:

```
KW   neuropathy; therapy; diagnosis; splice variant.
XX
OS   Homo sapiens.
XX
FH   Key              Location/Qualifiers
FT   Domain           1..360
FT                    /note= "extracellular domain, specifically claimed
FT                    in Claim 5(a)"
FT   Region           66..91
FT                    /note= "hydrophobic region"
FT   Region           101..284
FT                    /note= "mucin-like Ser/Thr-rich region, contains
FT                    sites for O-linked glycosylation"
FT   Domain           285..354
FT                    /note= "EGF-like domain"
FT   Domain           356..394
FT                    /note= "transmembrane domain"
XX
PN   WO902681-A1.
XX
PD   21-JAN-1999.
XX
PF   30-JUN-1998; 98WO-US13411.
XX
PR   24-JUL-1997; 97US-0899437.
PR   09-JUL-1997; 97US-0052019.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Godowski PJ, Mark MR, Zhang D;
XX
DR   WPI; 1999-120882/10.
DR   N-PSDB; AAX06989.
XX
PT   New isolated neuregulin related ligand-3 - used to develop products
PT   for treating nervous system disorders, e.g. stroke, ischaemia,
PT   infection, malignancy, Alzheimer's disease or Down's syndrome
XX
PS   Example 1; Page 78-81; 101pp; English.
XX
CC   This is the amino acid sequence of splice variant hNGR3B2 of human
CC   neuregulin related ligand NRG3, a novel member of the epidermal
CC   growth factor (EGF)-like family of protein ligands that binds to
CC   the ErbB4 receptor, but not to the ErbB2 or ErbB3 receptor, and
CC   which activates ErbB4 receptor tyrosine phosphorylation. The
CC   sequence was deduced from the nucleotide sequence of a cDNA clone
CC   (see AaX06989) from a foetal brain library. hNGR3B2 lacks amino
CC   acids 529-552 of hNGR3B1 (see AAW97618) but retains the EGF- like
CC   domain and is expected to exhibit biological activity. The invention
CC   provides human and murine NRG3 polypeptides (see AAW97617); expression
CC   vectors, host cells and methods for the recombinant production of
CC   NRG3s. The NRG3 polypeptides and polynucleotides and can be used to
CC   enhance the survival, proliferation or differentiation of cells
CC   having the ErbB4 receptor in vivo and in vitro. They can be used to
CC   prevent or treat damage to a nerve or damage to other NRG3-expressing
CC   or NRG3-responsive cells, e.g. brain, heart, or kidney cells. In
CC   particular, they can be used to treat diseases which involve neural
CC   cell growth such as demyelination, or damage or loss of glial cells
CC   nervous system has been damaged by e.g. trauma, surgery, stroke,
CC   (e.g. multiple sclerosis). They can be used to treat patients whose
CC   ischaemia, infection, metabolic disease, nutritional deficiency,
CC   malignancy, or toxic agents. NRG3 can also be used to treat
CC   motor neuron disorders such as amyotrophic lateral sclerosis (Lou
CC   Gehrig's disease), Bell's palsy, conditions involving spinal
CC   muscular atrophy or paralysis, neurodegenerative disorders such as
CC   Alzheimer's disease, Parkinson's disease, epilepsy, multiple
CC   sclerosis, Huntington's chorea, Down's syndrome, nerve deafness,
CC   and Meniere's disease. They can also be used to treat neuropathies
CC   associated with systemic disease including post-polio syndrome,
CC   hereditary neuropathies including Charcot-Marie-Tooth disease,
CC   Refsum's disease, abetalipoproteinemia, Tangier disease, Krabbe's
CC   disease, metachromatic leukodystrophy, Fabry's disease and
CC   Dejerine-Sottas syndrome, to treat disease of skeletal muscle of
```

Right column:

```
CC   smooth muscle, such as muscular dystrophy or diseases caused by
CC   skeletal or smooth muscle wasting. The products can also be used
CC   for detection, diagnosis, for the production of transgenic or
CC   knockout animals or for drug screening.
XX
SQ   Sequence    696 AA;
XX
Qy   1   HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   286 hfkpcrdkdlayclndgecfvietltgshkhcrckegyqgvrcdqfl 332

Query Match             100.0%;  Score 277;  DB 20;  Length 696;
Best Local Similarity   100.0%;  Pred. No. 2e-20;
Matches  47;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
RESULT  6
AAW97617
ID   AAW97617 standard; protein; 713 AA.
XX
AC   AAW97617;
XX
DT   10-MAY-1999  (first entry)
XX
DE   Mouse neuregulin related ligand NRG3.
XX
KW   Neuregulin related ligand; NRG3; mouse; ErbB4 receptor;
KW   signal transduction; nervous system disorder; neurodegeneration;
KW   neuropathy; therapy; diagnosis.
XX
OS   Mus sp.
XX
FH   Key              Location/Qualifiers
FT   Domain           1..362
FT                    /note= "extracellular domain, specifically claimed
FT                    in Claim 5(a)"
FT   Region           66..91
FT                    /note= "hydrophobic region"
FT   Region           105..286
FT                    /note= "mucin-like Ser/Thr-rich region, contains
FT                    sites for O-linked glycosylation"
FT   Domain           287..334
FT                    /note= "EGF-like domain"
FT   Domain           363..385
FT                    /note= "transmembrane domain"
XX
PN   WO902681-A1.
XX
PD   21-JAN-1999.
XX
PF   30-JUN-1998; 98WO-US13411.
XX
PR   24-JUL-1997; 97US-0899437.
PR   09-JUL-1997; 97US-0052019.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Godowski PJ, Mark MR, Zhang D;
XX
DR   WPI; 1999-120882/10.
DR   N-PSDB; AAX06987.
XX
PT   New isolated neuregulin related ligand-3 - used to develop products
PT   for treating nervous system disorders, e.g. stroke, ischaemia,
PT   infection, malignancy, Alzheimer's disease or Down's syndrome
XX
PS   Claim 5(b); Page 59-62; 101pp; English.
XX
CC   This is the amino acid sequence of murine neuregulin related ligand
CC   NRG3, a novel member of the epidermal growth factor (EGF)-like
CC   family of protein ligands that binds to the ErbB4 receptor, but not
CC   to the ErbB2 or ErbB3 receptor, and which activates ErbB4 receptor
```

CC   tyrosine phosphorylation. The sequence was deduced from the
CC   nucleotide sequences of cDNA clones (see AAX05987) from a mouse brain
CC   library. The EGF-like domain of NRG3 is distinct from those of NRG1
CC   or NRG2, and NRG3 displays receptor binding characteristics that are
CC   distinct from those of other neuregulins. The invention provides
CC   human and murine NRG3 polypeptides (see also AAW97618), expression
CC   vectors, host cells and methods for the recombinant production of
CC   NRG3s.
CC   The NRG3 polypeptides and polynucleotides and can be used to
CC   enhance the survival, proliferation or differentiation of cells
CC   having the ErbB4 receptor in vivo and in vitro. They can be used to
CC   prevent or treat damage to a nerve or damage to other NRG3-expressing
CC   or NRG3-responsive cells, e.g. brain, heart, or kidney cells. In
CC   particular, they can be used to treat diseases which involve neural
CC   cell growth such as demyelination, or damage or loss of glial cells
CC   (e.g. multiple sclerosis). They can be used to treat patients whose
CC   nervous system has been damaged by e.g. trauma, surgery, stroke,
CC   ischaemia, infection, metabolic disease, nutritional deficiency,
CC   malignancy, or toxic agents. NRG3 can also be used to treat
CC   motor neuron disorders such as amyotrophic lateral sclerosis (Lou
CC   Gehrig's disease), Bell's palsy, conditions involving spinal
CC   muscular atrophy or paralysis, neurodegenerative disorders such as
CC   Alzheimer's disease, Parkinson's disease, epilepsy, multiple
CC   sclerosis, Huntington's chorea, Down's syndrome, nerve deafness,
CC   and Meniere's disease. They can also be used to treat neuropathies
CC   associated with systemic disease including post-polio syndrome,
CC   hereditary neuropathies including Charcot-Marie-Tooth disease,
CC   Refsum's disease, abetalipoproteinemia, Tangier disease, Krabbe's
CC   disease, metachromatic leukodystrophy, Fabry's disease and
CC   Dejerine-Sottas syndrome, to treat disease of skeletal muscle of
CC   smooth muscle, such as muscular dystrophy or diseases caused by
CC   skeletal or smooth muscle wasting. The products can also be used
CC   for detection, diagnosis for the production of transgenic or
CC   knockout animals or for drug screening.
XX
SQ   Sequence   713 AA;

Query Match
Best Local Similarity   100.0%;   Score 277;   DB 20;   Length 713;
Matches   47;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;
                                                 Pred. No. 2e-20;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   288 hfkpcrdkdlayclndgecfvietltgshkhcrckegyqgvrcdqfl 334

RESULT 7
ID   AAY05452 standard; Protein; 720 AA.
XX
AC   AAY05452;
XX
DT   06-JUL-1999 (first entry)
XX
DE   Human heregulin-like factor sequence.
XX
KW   Human heregulin-like factor; HLF; cell growth regulator; diagnosis;
KW   neural system disorder; cancer.
XX
OS   Homo sapiens.
XX
XX
PN   WO9857989-A1.
XX
PD   23-DEC-1998.
XX
PF   16-JUN-1998; 98WO-US12403.
PR   17-JUN-1997; 97US-0049942.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (GEOU) UNIV GEORGETOWN.
XX
PI   Hijazi MM, King CR, Ruben SM, Young P;

XX
DR   WPI; 1999-095327/08.
XX
PT   New isolated heregulin-like factor - used to develop products for
PT   the diagnosis and treatment of disorders involving regulation of
PT   cell growth, particularly cancers
XX
PS   Disclosure; Page 97-99; 118pp; English.
XX
CC   This sequence is the human heregulin-like factor (HLF) of the
CC   invention. The HLF is involved in the regulation of cell growth.
CC   Detection of different levels of expression of the HLF gene can be used
CC   for the diagnosis of disorders, e.g. in the neural system. In
CC   particular, detection of different levels of HLF gene expression in cells
CC   or body fluid of an individual can be used for diagnosing cancer. The
CC   products can also be used in the treatment of disorders involving
CC   abnormal levels of HLF activity.
XX
SQ   Sequence   720 AA;

Query Match
Best Local Similarity   100.0%;   Score 277;   DB 20;   Length 720;
Matches   47;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;
                                                 Pred. No. 2.1e-20;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   286 hfkpcrdkdlayclndgecfvietltgshkhcrckegyqgvrcdqfl 332

RESULT 8
ID   AAW97618 standard; Protein; 720 AA.
XX
AC   AAW97618;
XX
DT   10-MAY-1999 (first entry)
XX
DE   Human neuregulin related ligand NRG3.
XX
KW   Neuregulin related ligand; NRG3; hNRG3B1; human; ErbB4 receptor;
KW   signal transduction; nervous system disorder; neurodegeneration;
KW   neuropathy; therapy; diagnosis.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FH
FT   Domain   1..360
FT   /note= "extracellular domain, specifically claimed
FT   in Claim 5(a)"
FT   Region   66..91
FT   /note= "hydrophobic region"
FT   Region   101..284
FT   /note= "mucin-like Ser/Thr-rich region, contains
FT   sites for O-linked glycosylation"
FT   Domain   285..354
FT   /note= "EGF-like domain"
FT   Domain   356..394
FT   /note= "transmembrane domain"
XX
OS   Homo sapiens.
XX
PN   WO9902681-A1.
XX
PD   21-JAN-1999.
XX
PF   30-JUN-1998; 98WO-US13411.
PR   24-JUL-1997; 97US-0899437.
PR   09-JUL-1997; 97US-0052019.
XX
PA   (GETH) GENENTECH INC.
XX
PI   Godowski PJ, Mark MR, Zhang D;

DR   WPI; 1999-120882/10.
DR   N-PSDB; AAX06988.
XX
PT   New isolated neuregulin related ligand-3 : used to develop products
PT   for treating nervous system disorders, e.g. stroke, ischaemia,
PT   Infection, malignancy, Alzheimer's disease or Down's syndrome
XX
PS   Claim 5(b); Page 66-69; 101pp; English.
XX
CC   This is the amino acid sequence of human neuregulin related ligand
CC   NRG3, a novel member of the epidermal growth factor (EGF)-like
CC   family of protein ligands that binds to the ErbB4 receptor, but not
CC   to the ErbB2 or ErbB3 receptor, and which activates ErbB4 receptor
CC   tyrosine phosphorylation. The sequence was deduced from the
CC   nucleotide sequence of a cDNA clone (see Aax06988) from a foetal brain
CC   library. The EGF-like domain of NRG3 is distinct from those of NRG1
CC   or NRG2, and NRG3 displays receptor binding characteristics that are
CC   distinct from those of other neuregulins. An alternatively spliced
CC   form of human NRG3 is provided in AAW97619. The invention provides
CC   human and murine NRG3 polypeptides (see also AAW97617), expression
CC   vectors, host cells and methods for the recombinant production of
CC   NRG3s. The NRG3 polypeptides and polynucleotides and can be used to
CC   enhance the survival, proliferation or differentiation of cells
CC   having the ErbB4 receptor in vivo and in vitro. They can be used to
CC   prevent or treat damage to a nerve or damage to other NRG3-expressing
CC   or NRG3-responsive cells, e.g. brain, heart, or kidney cells. In
CC   particular, they can be used to treat diseases which involve neural
CC   cell growth such as demyelination, or damage or loss of glial cells
CC   (e.g. multiple sclerosis). They can be used to treat patients whose
CC   nervous system has been damaged by e.g. trauma, surgery, stroke,
CC   ischaemia, infection, metabolic disease, nutritional deficiency,
CC   malignancy, or toxic agents. NRG3 can also be used to treat
CC   motor neuron disorders such as amyotrophic lateral sclerosis (Lou
CC   Gehrig's disease), Bell's palsy, conditions involving spinal
CC   muscular atrophy or paralysis, neurodegenerative disorders such as
CC   Alzheimer's disease, Parkinson's chorea, Down's syndrome, multiple
CC   sclerosis, Huntington's chorea, Parkinson's disease, nerve deafness,
CC   and Meniere's disease. They can also be used to treat neuropathies
CC   associated with systemic disease including post-polio syndrome,
CC   hereditary neuropathies including Charcot-Marie-Tooth disease,
CC   Refsum's disease, abetalippoproteinemia, Tangier disease, Krabbe's
CC   disease, metachromatic leukodystrophy, Fabry's disease and
CC   Dejerine-Sottas syndrome, to treat disease of skeletal muscle of
CC   smooth muscle, such as muscular dystrophy or diseases caused by
CC   skeletal or smooth muscle wasting. The products can also be used
CC   for detection, diagnosis, for the production of transgenic or
CC   knockout animals or for drug screening.
XX
SQ   Sequence 720 AA;
XX

Qy   Query Match   100.0%; Score 277; DB 20; Length 720;
     Best Local Similarity   100.0%;
     Matches   47; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
            |||||||||||||||||||||||||||||||||||||||||||||||
Db   286 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 332

RESULT 9
AAW05182
ID   AAW05182 standard; peptide; 52 AA.
XX
AC   AAW05182;
XX
DT   04-JUN-1997 (first entry)
XX
DE   Neu differentiation factor/heregulin-alpha/beta form EGF-like domain.
XX
KW   neu differentiation factor; heregulin; epidermal growth factor;
KW   EGF; colon epithelial cell proliferation; Schwann cell; nerve;
KW   damage; colitis; ulcer.
XX
OS   Synthetic.
XX
PN   WO9631599-A1.
XX
PD   10-OCT-1996.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   27-MAR-1996; 96WO-US04262.
XX
PF   Cranahan JF, Hara S, Lu HS, Mayer JP, Yoshinaga SK;
XX
PR   06-APR-1995; 95US-0417640.
XX
DR   WPI; 1996-465022/46.
XX
PS   Claim 1; Page 27; 37pp; English.
XX
PT   Peptide(s) derived from neu differentiation factor/heregulin
PT   proteins - specifically from epidermal growth factor-like domain,
PT   stimulate proliferation of colon epithelial cells and Schwann cells
XX
CC   The peptides AAW05182-W05185 are based on neu differentiation factor
CC   (NDF)/heregulin alpha and beta form EGF-like domains in various
CC   combinations. The peptides maintain the survival and proliferation of
CC   Schwann cells and cause proliferation, growth and differentiation of
CC   colon epithelial cells. According to the invention, they are useful to treat
CC   (in vivo) a disease or disorder of the colon (e.g. colitis or an
CC   ulcer) or of the nervous system (e.g. nerve damage caused by trauma).
XX
SQ   Sequence 52 AA;
XX

Qy   Query Match   42.1%; Score 116.5; DB 17; Length 52;
     Best Local Similarity   33.3%;
     Matches   16; Conservative   15; Mismatches   16; Indels   1; Gaps   1;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL 47
            | ||: ||||| ||||: : || |:|| || | |:|||||||||| :
Db   2 HLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCKCQPGFTGARCQNYV 49

RESULT 10
AAB12602
ID   AAB12602 standard; Peptide; 52 AA.
XX
AC   AAB12602;
XX
DT   09-NOV-2000 (first entry)
XX
DE   Human NDF EGF-like domain derived peptide SEQ ID NO:1.
XX
KW   Human; sensory epithelial cell; growth; stimulant; inner ear; EGF;
KW   epithelial growth factor; NDF; heregulin; monoclonal antibody;
KW   adult rat utricular epithelium.
XX
OS   Homo sapiens.
XX
PN   US6080845-A.
XX
PD   27-JUN-2000.
XX
PF   28-JAN-1999; 99US-0238182.
XX
PR   05-AUG-1998; 98US-0129549.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Carnahan JF;
XX
DR   WPI; 2000-451229/39.
DR   N-PSDB; AAA60998.

XX
PT   Novel monoclonal antibody against adult rat utricular epithelium useful
PT   in study and research with such organs and tissue -
XX
PS
XX
    Example 1; Fig 1; 12pp; English.
XX
CC   The present invention describes a monoclonal antibody (I) deposited with
CC   the American Type Culture Collection under accession number HB-12598.
CC   The antibodies are used in study and research with adult rat utricular
CC   organs and tissues. The present sequence is a peptide derived from
DE   the epithelial growth factor (EGF) like domain of NDF-alpha and
CC   NDF-beta (members of the NDF/heregulin protein family, which is used in
CC   the exemplification of the present invention. The peptide acts as a
CC   growth stimulant for sensory epithelial cells of the inner ear.
XX
SQ   Sequence   52 AA;

RESULT 11
ID   AAY69983
AC   AAY69983;
XX
DT   20-APR-2000   (first entry)
XX
DE   NDF/heregulin protein family derivative peptide.
XX
KW   NDF/heregulin protein family; growth factor; proliferation stimulator;
KW   sensory epithelial cell; inner ear; vestibular disorder; hearing loss;
KW   balance disorder; therapy.
XX
OS   Synthetic.
XX
PN   US6017886-A.
XX
PD   25-JAN-2000.
XX
PF   23-FEB-1999;   99US-0255974.
PR   05-AUG-1998;   98US-0129549.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Carnahan JF;
XX
DR   WPI; 2000-136686/12.
XX
PT   Stimulating the proliferation of sensory epithelial cells of the inner
PT   ear with a peptide, useful for treating hearing loss -
XX
PS   Claim 1; Fig 1; 11pp; English.
XX
CC   This sequence represents an NDF/heregulin protein family derivative
CC   peptide. The invention relates to a method for stimulating the
CC   proliferation of sensory epithelial cells of the inner ear, comprising
CC   contacting the cells with this sequence. The peptide acts as a growth
CC   factor to stimulate proliferation of cells in the sensory epithelium. The
CC   method is useful for the treatment of vestibular disorders (i.e. a
CC   balance disorder) or for treating hearing loss.
XX
SQ   Sequence   52 AA;

Qy   Query Match   42.1%;   Score 116.5;   DB 21;   Length 52;
      Best Local Similarity 33.3%;   Pred. No. 3.1e-05;
      Matches 16;   Conservative 15;   Mismatches 16;   Indels 1;   Gaps 1;

Qy   1   HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL 47
              | :|: :|:| ||||:||: :: |:|: |: : || :| ::
Db   2   hlvkcaekektfcvnggecfmvkdlsnpsrylckcqpgftgarcqnyv 49

RESULT 12
ID   AAW05184
AC   AAW05184;
XX
DT   04-JUN-1997   (first entry)
XX
DE   Neu differentiation factor/heregulin-alpha form EGF-like domain.
XX
KW   NDF; neu differentiation factor; heregulin; epidermal growth factor;
KW   EGF; colon epithelial cell proliferation; Schwann cell; nerve;
KW   damage; colitis; ulcer.
XX
OS   Synthetic.
XX
PN   WO9631599-A1.
XX
PD   10-OCT-1996.
XX
PF   27-MAR-1996;   96WO-US04262.
XX
PR   06-APR-1995;   95US-0417640.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Cranahan JF, Hara S, Lu HS, Mayer JP, Yoshinaga SK;
XX
DR   WPI; 1996-465022/46.
XX
PT   Peptide(s) derived from neu differentiation factor/heregulin
PT   proteins - specifically from epidermal growth factor-like domain,
PT   stimulate proliferation of colon epithelial cells and Schwann cells
XX
PS   Disclosure; Page 24; 37pp; English.
XX
CC   The peptides AAW05182-W05185 are based on neu differentiation factor
CC   (NDF)/heregulin alpha and beta form EGF-like domains in various
CC   combinations. The peptides maintain the survival and proliferation of
CC   Schwann cells and cause proliferation, growth and differentiation of
CC   colon epithelial cells. Accordingly, they are useful to treat (in vitro
CC   or in vivo) a disease or disorder of the colon (e.g. colitis or an
CC   ulcer) or of the nervous system (e.g. nerve damage caused by trauma).
XX
SQ   Sequence   52 AA;

Qy   Query Match   42.1%;   Score 116.5;   DB 21;   Length 52;
      Best Local Similarity 33.3%;   Pred. No. 3.1e-05;
      Matches 16;   Conservative 15;   Mismatches 16;   Indels 1;   Gaps 1;

Qy   1   HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL 47
              | :|: :|:| ||||:||: :: |:|: |: : || :| ::
Db   2   hlvkcaekektfcvnggecfmvkdlsnpsrylckcqpgftgarcqnyv 49

RESULT 13
ID   AAR46918
AC   AAR46918;
XX
DT   28-JUL-1994   (first entry)
XX
DE   AAR46918 standard; Protein; 63 AA.

Qy   Query Match   41.0%;   Score 113.5;   DB 17;   Length 52;
      Best Local Similarity 34.8%;   Pred. No. 6.3e-05;
      Matches 16;   Conservative 14;   Mismatches 15;   Indels 1;   Gaps 1;

Qy   1   HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQ 45
              | :|: :|:| ||||:||: :: |:|: |: : || :| :
Db   2   hlvkcaekektfcvnggecfmvkdlsnpsrylckcqpgftgarcte 47

```
DE   EGFL2.
XX
KW   Glial growth factor; GGF; heregulin; mitogenesis;
KW   Schwann cell; tumour; central nervous system; erbB2 receptor;
KW   antiproliferative; epidermal growth factor; EGF.
XX
PN   WO9403644-A.
XX
PD   17-FEB-1994.
XX
PF   10-AUG-1993;
XX
PR   10-AUG-1992; 92US-0927337.
PR   25-SEP-1992; 92US-0951747.
PR   01-DEC-1992; 92US-0984085.
PR   29-JAN-1993; 93US-0011396.
XX
PA   (CAMB-) CAMBRIDGE NEUROSCIENCE INC.
XX
PI   Gwynne DI, Marchionni M, McBurney RN;
XX
DR   WPI; 1994-0657931/08.
DR   N-PSDB; AAQ58324.
XX
PT   Glial growth factor DNA encoding numerous polypeptide factors
PT   used for inhibiting cell proliferation - for treating carcinoma
PT   and nervous disorders
XX
PS   Disclosure; Fig 40; 178pp; English.
XX
CC   The GGF coding segments include regions with EGF-like homology.
CC   These EGF-like domains can be required for the activation of
CC   mitogenesis in the binding reaction between GGF ligands contg.
CC   such domains and the erbB2 receptor. Pref. antiproliferative
CC   factors are those which lack these EGF-like domains.
XX
SQ   Sequence   63 AA;

Query Match      41.0%;  Score 113.5;  DB 15;  Length 63;
Best Local Similarity   34.8%;  Pred. No. 7.5e-05;
Matches  16;  Conservative  14;  Mismatches  15;  Indels  1;  Gaps  1;

Qy     1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQ 45
           | :|: :|:| ||||:|||:::|:||::||| :|||:| ||| : ||
Db     2 hlvkcaekektfcvnggecfmvkdisnpsrylckcqpfgarcte 47
```

```
RESULT 14
AAR55659
ID   AAR55659 standard; Protein; 63 AA.
XX
AC   AAR55659;
XX
DT   28-JUL-1994 (first entry)
XX
DE   EGFL2.
XX
KW   Glial growth factor; GGF; heregulin; mitogenesis;
KW   Schwann cell; tumour; central nervous system;
KW   epidermal growth factor; EGF.
XX
PN   WO9400140-A.
XX
PD   06-JAN-1994.
XX
PF   29-JUN-1993;
XX
PR   30-JUN-1992; 92US-0907138.
PR   03-SEP-1992; 92US-0940389.
PR   23-OCT-1992; 92US-0965173.
PR   24-MAR-1993; 93US-0036555.
XX
PA   (CAMB-) CAMBRIDGE NEUROSCIENCE.
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Chen MS, Goodearl A, Hiles I, Marchionni M, Minghetti L;
PI   Stroobant P, Waterfield M;
XX
DR   WPI; 1994-025882/03.
DR   N-PSDB; AAQ62843.
XX
PS   Claim 53; Fig 39; 178pp; English.
XX
PT   Glial mitogenic polypeptide factors - useful for stimulating
PT   glial cell mitogenesis and treating glial cell tumours
XX
CC   EGFL1, EGFL2, EGFL3, EGFL4, EGFL5 and EGFL6 are used
CC   for the stimulation of glial cell mitogenesis in vivo
CC   and in vitro.
XX
SQ   Sequence   63 AA;

Query Match      41.0%;  Score 113.5;  DB 15;  Length 63;
Best Local Similarity   34.8%;  Pred. No. 7.5e-05;
Matches  16;  Conservative  14;  Mismatches  15;  Indels  1;  Gaps  1;

Qy     1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQ 45
           | :|: :|:| ||||:|||:::|:||::||| :|||:| ||| : ||
Db     2 hlvkcaekektfcvnggecfmvkdisnpsrylckcqpfgarcte 47
```

```
RESULT 15
AAR67250
ID   AAR67250 standard; Protein; 63 AA.
XX
AC   AAR67250;
XX
DT   15-AUG-1995 (first entry)
XX
DE   Human epidermal like growth factor 2 (EGFL2).
XX
KW   Epidermal like growth factor 2; mammalian muscle cell treatment;
KW   skeletal; cardiac; smooth; acetylcholine receptor deficiency;
XX
OS   Homo sapiens.
XX
PN   WO9426298-A.
XX
PD   24-NOV-1994.
XX
PF   06-MAY-1994;
XX
PR   06-MAY-1993; 93US-0059022.
PR   08-MAR-1994; 94US-0209204.
XX
PA   (CAMB-) CAMBRIDGE NEUROSCIENCE.
XX
PI   Gwynne DI, Marchionni M, Sklar R;
XX
DR   WPI; 1995-006353/01.
DR   N-PSDB; AAQ74915.
XX
PS   Claim 34; Pages 148-149; 24pp; English.
XX
PT   Treating mammalian muscle diseases and disorders - by admin. of
PT   GGF2 and other specified polypeptide(s) which bind the p185erbB2
PT   receptor.
XX
CC   AAQ74915 encodes AAR67250 human epidermal like growth factor 2 (EGFL2).
CC   The glial cell mitogenic activity of EGFL2 can be used to treat a
CC   variety of mammalian skeletal, cardiac and smooth muscle diseases,
CC   including acetylcholine receptor deficiency.
XX
```

SQ   Sequence   63 AA;

Query Match   41.0%;   Score 113.5;   DB 16;   Length 63;
Best Local Similarity 34.8%;   Pred. No. 7.5e-05;
Matches   16;   Conservative   14;   Mismatches   15;   Indels   1;   Gaps   1;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQ 45
             | |:| :|:|| |||||:::|::| :: | | |:| | :| :
Db   2 hivkcaektfcvnggecfmvkdlsnpsrylckcqpgftgarcte 47

Search completed: February 7, 2002, 12:59:51
Job time: 305 sec

```
                               GenCore version 4.5
                      Copyright (c) 1993 - 2000 Compugen Ltd.

QM protein - protein search, using sw model

Run on:          February 7, 2002, 13:00:12 ; Search time 10.15 Seconds
                 (without alignments)
                 169.778 Million cell updates/sec

Title:           US-09-480-977-4
Perfect score:   277
Sequence:        1 HFKPGRDKDLAYCLNDGECF.........SHKHCRCKEGYQGVRCDFL 47

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters:   100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_39:*

   Pred. No. is the number of results predicted by chance to have a
   score greater than or equal to the score of the result being printed,
   and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	Match	DB ID	Description
1	277	100.0	713	1	NRG3_MOUSE O35181	mus musculu
2	277	100.0	720	1	NRG3_HUMAN P56975	homo sapien
3	126.5	45.7	677	1	NRG1_XENLA O93383	xenopus lae
4	113.5	41.0	639	1	NRG1_HUMAN Q02297	h.pro-neure
5	110	39.7	115	1	NRG4_MOUSE Q9wtx4	mus musculu
6	104.5	37.7	296	1	SMDF_HUMAN Q15491	homo sapien
7	104	37.5	756	1	NRG2_MOUSE P56974	mus musculu
8	102.5	37.0	662	1	NRG1_RAT Q43322	r pro-neure
9	98.5	35.6	602	1	NRG1_CHICK O05199	gallus gall
10	90	32.5	850	1	NRG1_HUMAN O14511	homo sapien
11	90	32.5	868	1	NRG2_RAT O35569	rattus norv
12	89.5	32.3	169	1	EREG_HUMAN O14944	homo sapien
13	85	30.7	80	1	GRFA_SFVKA P08441	shope fibro
14	83	30.0	2531	1	NTC1_MOUSE O01705	mus musculu
15	82.5	29.8	177	1	BTC_MOUSE Q05928	mus musculu
16	82.5	29.8	178	1	BTC_BOVIN Q9ttc5	bos taurus
17	82	29.6	85	1	GRFA_MYXVL P08072	myxoma viru
18	82	29.6	230	1	SPIT_DROME O01083	drosophila
19	80.5	29.1	178	1	BTC_HUMAN P35070	homo sapien
20	79.5	28.7	140	1	GRFA_VACCV P01136	vaccinia vi
21	79.5	28.7	142	1	GRFA_VACCC P20494	vaccinia vi
22	79.5	28.7	159	1	TGFA_MOUSE P48030	mus musculu
23	79.5	28.7	159	1	TGFA_RAT P01134	rattus norv
24	79	28.5	2531	1	NTC1_RAT O07008	rattus norv
25	79	28.5	1207	1	EGF_HUMAN P01133	homo sapien
26	77.5	28.0	714	1	DLL1_RAT P97677	rattus norv
27	77.5	28.0	722	1	DLL1_MOUSE Q61483	mus musculu
28	76	27.4	484	1	LEM2_PIG P98110	sus scrofa
29	76	27.4	1217	1	EGF_MOUSE P01132	mus musculu
30	76	27.4	2139	1	CRB_DROME P10040	drosophila
31	75	27.1	611	1	LEM2_CANFA P33730	canis famil
32	74.5	26.9	603	1	FA12_CAVPO Q04962	cavia porce
33	74.5	26.9	723	1	DLL1_HUMAN O00548	homo sapien
34	74.5	26.9	1429	1	LI12_CAEEL P45585	caenorhabd1
35	74	26.7	551	1	LEM2_RABIT P27113	oryctolagus
36	74	26.7	610	1	LEM2_HUMAN P16581	homo sapien
37	74	26.7	2871	1	FBN1_BOVIN P98133	bos taurus
38	74	26.7	2871	1	FBN1_HUMAN P35555	homo sapien
39	74	26.7	2871	1	FBN1_MOUSE Q61554	mus musculu
40	74	26.7	294	1	GRK_DROME Q9tv36	drosophila
41	74	26.7	409	1	FBN1_PIG P42287	sus scrofa
42	73	26.4	409	1	MRGM_PIG P79385	sus scrofa
43	73	26.4	1964	1	NTC4_MOUSE P31695	mus musculu
44	73	26.4	2524	1	NOTC_XENLA P21783	xenopus lae
45	72.5	26.2	121	1	TGFA_MACMU P55244	macaca mula

ALIGNMENTS

```
RESULT 1
NRG3_MOUSE
ID   NRG3_MOUSE     STANDARD;      PRT;   713 AA.
AC   O35181;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PRO-NEUREGULIN-3 PRECURSOR (PRO-NRG3) [CONTAINS: NEUREGULIN-3 (NRG-
DE   3)].
GN   NRG3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97420720; PubMed=9275162;
RA   Zhang D., Sliwkowski M.X., Mark M., Frantz G., Akita R., Sun Y.,
RA   Hillan K., Crowley C., Brush J., Godowski P.J.;
RT   "Neuregulin-3 (NRG3): a novel neural tissue-enriched protein that
RT   binds and activates ErbB4.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:9562-9567(1997).
CC   -!- FUNCTION: DIRECT LIGAND FOR THE ERBB4 TYROSINE KINASE RECEPTOR.
CC       BINDING RESULTS IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC       ACTIVATION OF THE RECEPTOR. DOES NOT BIND TO THE EGF RECEPTOR,
CC       ERBB2 OR ERBB3 RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
CC       A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
CC       MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN SYMPATHETIC, MOTOR, AND SENSORY
CC       NEURONS.
CC   -!- DEVELOPMENTAL STAGE: DETECTED AS EARLY AS E11. IN E13 EMBRYOS,
CC       DETECTED MAINLY IN THE NERVOUS SYSTEM. IN E16 EMBRYOS, DETECTED IN
CC       THE BRAIN, SPINAL CORD, TRIGEMINAL, VESTIBULAR-COCHLEAR, AND
CC       SPINAL GANGLIA. IN ADULTS, EXPRESSED IN SPINAL CORD, AND NUMEROUS
CC       BRAIN REGIONS.
CC   -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC       OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC       PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC       DIMERIZATION (BY SIMILARITY).
CC   -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC       DOMAIN (BY SIMILARITY).
CC   -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
CC       EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC       FORM (BY SIMILARITY).
CC   -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AF010130; AAB70914.1; -.
DR MGD; MGI:1097165; Nrg3.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002154; Neuregulin.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF02158; Neuregulin; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
KW Growth factor; EGF-like domain; Transmembrane; Multigene family.
FT CHAIN   1   713   PRO-NEUREGULIN-3.
FT CHAIN   1   362   NEUREGULIN-3.
FT DOMAIN   1   361   EXTRACELLULAR (POTENTIAL).
FT TRANSMEM   363   383   INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT DOMAIN   384   713   CYTOPLASMIC (POTENTIAL).
FT DOMAIN   105   287   SER/THR-RICH.
FT DOMAIN   288   331   EGF-LIKE.
FT DOMAIN   13   21   POLY-ALA.
FT DOMAIN   26   34   POLY-ALA.
FT DOMAIN   127   135   POLY-THR.
FT DOMAIN   250   253   POLY-ALA.
FT DOMAIN   254   263   POLY-SER.
FT DOMAIN   264   267   POLY-THR.
FT DISULFID   292   306   BY SIMILARITY.
FT DISULFID   300   319   BY SIMILARITY.
FT DISULFID   321   330   BY SIMILARITY.
SQ SEQUENCE   713 AA;   77369 MW; 9F7D1D5E7FC8DCF0 CRC64;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   288 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 334

Query Match   100.0%;   Score 277; DB 1; Length 713;
Best Local Similarity   100.0%;   Pred. No. 6.2e-26;
Matches   47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2
ID   NRG3_HUMAN   STANDARD;   PRT;   720 AA.
AC   P56975;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PRO-NEUREGULIN-3 PRECURSOR (PRO-NRG3) [CONTAINS: NEUREGULIN-3 (NRG-
DE   3)].
GN   NRG3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=97420720; PubMed=9275162;
RA   Zhang D., Sliwkowski M.X., Mark M., Frantz G., Akita R., Sun Y.,
RA   Hillan K., Crowley C., Brush J., Godowski P.J.;
RT   "Neuregulin-3 (NRG3): a novel neural tissue-enriched protein that
RT   binds and activates ErbB4.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:9562-9567(1997).
CC -!- FUNCTION: DIRECT LIGAND FOR THE ERBB4 TYROSINE KINASE RECEPTOR.
CC   BINDING RESULTS IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC   ACTIVATION OF THE RECEPTOR. DOES NOT BIND TO THE EGF RECEPTOR,
CC   ERBB2 OR ERBB3 RECEPTORS.
CC -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
CC   A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
CC   MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN MOST REGIONS OF THE BRAIN
CC   EXPRESSED AT LOWER LEVEL IN
CC   WITH THE EXCEPTION OF CORPUS CALLOSUM.

CC   TESTIS. NOT DETECTED IN HEART, PLACENTA, LUNG, LIVER, SKELETAL
CC   MUSCLE, KIDNEY, PANCREAS, SPLEEN, THYMUS, PROSTATE, OVARY, SMALL
CC   INTESTINE, COLON AND PERIPHERAL BLOOD LEUKOCYTES.
CC -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC   OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC   PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC   DIMERIZATION (BY SIMILARITY).
CC -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC   DOMAIN (BY SIMILARITY).
CC -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
CC   EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC   FORM (BY SIMILARITY).
CC -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
CC   SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002154; Neuregulin.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF02158; Neuregulin; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
KW Growth factor; EGF-like domain; Transmembrane; Multigene family.
FT CHAIN   1   720   PRO-NEUREGULIN-3.
FT CHAIN   1   359   NEUREGULIN-3.
FT DOMAIN   1   360   EXTRACELLULAR (POTENTIAL).
FT TRANSMEM   361   381   INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT DOMAIN   382   720   CYTOPLASMIC (POTENTIAL).
FT DOMAIN   105   285   SER/THR-RICH.
FT DOMAIN   286   329   EGF-LIKE.
FT DOMAIN   5   8   POLY-ALA.
FT DOMAIN   13   21   POLY-ALA.
FT DOMAIN   26   34   POLY-ALA.
FT DOMAIN   127   135   POLY-THR.
FT DISULFID   290   304   BY SIMILARITY.
FT DISULFID   298   317   BY SIMILARITY.
FT DISULFID   319   328   BY SIMILARITY.
SQ SEQUENCE   720 AA;   77900 MW; A4D6F10DB95A693 CRC64;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 47
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   286 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQFL 332

Query Match   100.0%;   Score 277; DB 1; Length 720;
Best Local Similarity   100.0%;   Pred. No. 6.2e-26;
Matches   47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3
ID   NRG1_XENLA   STANDARD;   PRT;   677 AA.
AC   O93383; Q9W6N0;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PRO-NEUREGULIN-1 PRECURSOR (PRO-NRG1) [CONTAINS: NEUREGULIN-1].
GN   NRG1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA1, AND ALTERNATIVE SPLICING.
RX   MEDLINE=98352126; PubMed=9685585;
RA   Yang J.F., Zhou H., Pun S., Ip N.Y., Peng H.B., Tsim K.W.K.;
RT   "Cloning of cDNAs encoding xenopus neuregulin: expression in myotomal
RT   muscle during embryo development.";
RL   Brain Res. Mol. Brain Res. 58:59-73(1998).

RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM CRD).
RX   MEDLINE=99316087; PubMed=10383827;
RA   Yang J.F., Zhou H., Choi R.C., Ip N.Y., Peng H.B., Tsim K.W.K.;
RL   "A cysteine-rich form of Xenopus neuregulin induces the expression of
RL   acetylcholine receptors in cultured myotubes.";
RL   Mol. Cell. Neurosci. 13:415-429(1999).
CC   -!- FUNCTION: DIRECT LIGAND FOR THE ERBB TYROSINE KINASE RECEPTORS.
CC   INDUCES EXPRESSION OF ACETYLCHOLINE RECEPTOR IN SYNAPTIC NUCLEI.
CC   -!- SUBCELLULAR LOCATION: EXISTS AS A TYPE I MEMBRANE PROTEIN AND AS A
CC   PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE MEMBRANE-
CC   BOUND FORM DOES NOT SEEM TO BE ACTIVE. (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; ALPHA1 (SHOWN HERE) AND
CC   CRD/CRD-NRG1 (CYSTEINE RICH DOMAIN); ARE PRODUCED BY ALTERNATIVE
CC   SPLICING. ISOFORMS HAVE ALPHA OR BETA-TYPE EGF-LIKE DOMAINS.
CC   -!- TISSUE SPECIFICITY: ISOFORM ALPHA1 IS EXPRESSED IN BRAIN AND
CC   MUSCLE. ISOFORM CRD IS EXPRESSED IN BRAIN AND SPINAL CORD, BUT
CC   AT VERY LOW LEVEL IN MUSCLE.
CC   -!- DEVELOPMENTAL STAGE: STRONG EXPRESSION IN DEVELOPING BRAIN AND
CC   SPINAL CORD OF THE EMBRYO. ALSO EXPRESSED IN THE MYOTOMAL MUSCLE.
CC   -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC   OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC   PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC   DIMERIZATION (BY SIMILARITY).
CC   -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC   DOMAIN.
CC   -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
CC   EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC   FORM.
CC   -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
CC   SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
CC   ---

CC   ---
DR   EMBL; AF076618; AAC26804.1; -.
DR   EMBL; AF142632; AAD33893.1; -.
DR   HSSP; Q02297; 1HRF.
DR   InterPro; IPR005561; EGF-like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR002154; Neuregulin.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00047; Ig; 1.
DR   Pfam; PF02158; Neuregulin; 1.
DR   PRINTS; PR01089; NEUREGULIN.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW   Transmembrane; Alternative splicing.
FT   CHAIN   1   259   NEUREGULIN ALPHA1 (BY SIMILARITY).
FT   CHAIN   1   677   PRO-NEUREGULIN ALPHA1; MEMBRANE-BOUND
FT   FORM (BY SIMILARITY).
FT   DOMAIN   1   260   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   261   280   INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT   DOMAIN   281   677   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN   50   119   IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN   188   232   EGF-LIKE.
FT   DISULFID   57   116   BY SIMILARITY.
FT   DISULFID   192   206   BY SIMILARITY.
FT   DISULFID   200   220   BY SIMILARITY.
FT   DISULFID   222   231   BY SIMILARITY.
FT   DOMAIN   1   25   LYS-RICH.

FT   CARBOHYD   124   124   N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   130   130   N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1   136   IKFQSVQEGKKLVLKCQAVSEQPSLKFRMRFGKEXIGARNK
FT   PDSKPEHIKIRKRKKSSELQISKASSANGEYXCMVSNQLG
FT   NDTVTVNVTIVPK -> MSEDTAEGLAONQCSEQSSDPSAE
FT   LQNEESMPETQDEEETTNGITGLAITCCVCLEADRLRICLN
FT   SEKICIPILACISISLCICIAGLKWVFVDKIFEYDSPPHLD
FT   PGHRGODLILYDTPAPSTLVPSSVRTLPVIIPTDSKAAVT
FT   FKFGTSLLPTE (IN ISOFORM CRD).
FT   VARSPLIC   223   252   KFGFTGARCTEDPLRVVRSEKHLGIEFME -> PNEFTGD
FT   RCONYVWASFYK (IN ISOFORM CRD).
SQ   SEQUENCE   677 AA; 75794 MW; 49279EBF5BAE396F CRC64;

Qy   1 HKKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQ 45
        | | || | || || | |:: | | |: :| |:| |||| | ||
Db 188 HLIKCSDKEKTYCVNGGECYVLNGITSSNQFMCKCKPGFTGARCTE 233

Query Match   45.7%; Score 126.5; DB 1; Length 677;
Best Local Similarity   43.5%; Pred. No. 6.4e-08;
Matches   20; Conservative   10; Mismatches   15; Indels   1; Gaps   1;

RESULT 4
NRG1_HUMAN
ID   NRG1_HUMAN   STANDARD;   PRT;   639 AA.
AC   Q02297; Q02298; Q02299; Q12779; Q12780; Q12781; Q12782; Q12783;
AC   Q12784; Q07110; Q07111; Q90PE3; O14667;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PRO-NEUREGULIN-1 PRECURSOR (PRO-NRG1) [CONTAINS: NEUREGULIN-1 (NEU
DE   DIFFERENTIATION FACTOR) (HEREGULIN) (HRG) (BREAST CANCER CELL
DE   DIFFERENTIATION FACTOR P45) (ACETYLCHOLINE RECEPTOR INDUCING ACTIVITY)
DE   (ARIA) (SENSORY AND MOTOR NEURON-DERIVED FACTOR) (GLIAL GROWTH
DE   FACTOR)].
GN   NRG1 OR HGL OR NDF OR HRGA OR GGF OR SMDF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ALPHA; BETA1; BETA2; BETA3), AND PARTIAL SEQUENCE.
RC   TISSUE=Pituitary, and kidney adenocarcinoma;
RX   MEDLINE=92271253; PubMed=1350381;
RA   Holmes W.E., Sliwkowski M.X., Akita R.W., Henzel W.J., Lee J.,
RA   Park J.W., Yansura D., Abadi N., Raab H., Lewis G.D., Shepard H.M.,
RA   Kuang W.-J., Wood W.I., Goeddel D.V., Vandlen R.L.;
RL   "Identification of heregulin, a specific activator of p185erbB2.";
RL   Science 256:1205-1210(1992).
RN   [2]
RP   SEQUENCE FROM N.A. (ALPHA1; ALPHA2B; ALPHA3; BETA1A; BETA2 & BETA3).
RC   TISSUE=Pituitary, and kidney adenocarcinoma;
RX   MEDLINE=94158863; PubMed=7509448;
RA   Wen D., Suggs S.V., Karunagaran D., Liu N., Cupples R.L., Luo Y.,
RA   Janssen A.M., Ben-Baruch N., Trollinger D.B., Jacobsen V.L.,
RA   Meng S.-Y., Lu H.S., Hu S., Chang D., Yang W., Yanigahara D.,
RA   Koski R.A., Yarden Y.;
RL   "Structural and functional aspects of the multiplicity of Neu
RL   differentiation factors.";
RL   Mol. Cell. Biol. 14:1909-1919(1994).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA).
RX   MEDLINE=92209945; PubMed=1348215;
RA   Peles E., Bacus S.S., Koski R.A., Lu H.S., Wen D., Ogden S.G.,
RA   Levy R.B., Yarden Y.;
RL   "Isolation of the neu/HER-2 stimulatory ligand: a 44 kd glycoprotein
RL   that induces differentiation of mammary tumor cells.";
RL   Cell 69:205-216(1992).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORMS BETA3 AND GGF2).
RC   TISSUE=Brain;
RX   MEDLINE=93205115; PubMed=8096067;

RA   Marchionni M.A., Goodearl A.D.J., Chen M.S., Bermingham-McDonogh O.,
RA   Kirk C., Hendricks M., Danehy F., Misumi D., Sudhalter J.,
RA   Kobayashi K., Wroblewski D., Lynch C., Baldassarre M., Hiles I.,
RA   Davis J.B., Hsuan J.J., Totty N.F., Otsu M., McBurney R.N.,
RA   Waterfield M.D., Stroobant P., Gwynne D.;
RT   "Glial growth factors are alternatively spliced erbB2 ligands
RT   expressed in the nervous system.";
RL   Nature 362:312-318(1993).
RN   [5]
RP   SEQUENCE FROM N.A. OF GAMMA-HEREGULIN FUSION PROTEIN.
RC   TISSUE=Breast cancer;
RX   MEDLINE=97472144; PubMed=9333014;
RA   Schaefer G., Fitzpatrick V.D., Sliwkowski M.X.;
RT   "Gamma-heregulin: a novel heregulin isoform that is an autocrine
RT   growth factor for the human breast cancer cell line, MDA-MB-175.";
RL   Oncogene 15:1385-1394(1997).
RN   [6]
RP   SEQUENCE OF 1-210 FROM N.A.
RA   Schoumacher F., Herzer S., Flury N., Kueng W., Mueller H.,
RA   Eppenberger U.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 19-27.
RX   MEDLINE=93667031; PubMed=7689552;
RA   Culouscou J.-M., Plowman G.D., Carlton G.W., Green J.M., Shoyab M.;
RT   "Characterization of a breast cancer cell differentiation factor that
RT   specifically activates the HER4/p180erbB4 receptor.";
RL   J. Biol. Chem. 268:18407-18410(1993).
RN   [8]
RP   CHROMOSOMAL TRANSLOCATION.
RX   MEDLINE=99455251; PubMed=10523851;
RA   Wang X.-Z., Jolicoeur E.M., Conte N., Chaffanet M., Zhang Y.,
RA   Mozziconacci M.-J., Felner H., Birnbaum D., Pebusque M.-J., Ron D.;
RT   "Gamma-heregulin is the product of a chromosomal translocation fusing
RT   the DOC4 and HGL/NRG1 genes in the MDA-MB-175 breast cancer
RT   line.";
RL   Oncogene 18:5718-5721(1999).
RN   [9]
RP   CHROMOSOMAL TRANSLOCATION.
RX   MEDLINE=20065180; PubMed=10597312;
RA   Liu X., Baker E., Eyre H.J., Sutherland G.R., Zhou M.;
RT   "Gamma-heregulin: a fusion gene of DOC-4 and neuregulin-1 derived from
RT   a chromosome translocation.";
RL   Oncogene 18:7110-7114(1999).
RN   [10]
RP   STRUCTURE BY NMR OF 175-241 (ISOFORM ALPHA).
RX   MEDLINE=94341264; PubMed=8062828;
RA   Nagata K., Kohda D., Hatanaka H., Ichikawa S., Matsuda S.,
RA   Yamamoto T., Suzuki A., Inagaki F.;
RT   "Solution structure of the epidermal growth factor-like domain of
RT   heregulin-alpha, a ligand for p180erbB-4.";
RL   EMBO J. 13:3517-3523(1994).
CC   -!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE
CC   RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
CC   RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC   ACTIVATION OF THE ERBB RECEPTORS. THE MULTIPLE ISOFORMS PERFORM
CC   DIVERSE FUNCTIONS SUCH AS INDUCING GROWTH AND DIFFERENTIATION OF
CC   EPITHELIAL, GLIAL, NEURONAL, AND SKELETAL MUSCLE CELLS; INDUCING
CC   EXPRESSION OF ACETYLCHOLINE RECEPTOR IN SYNAPTIC VESICLES DURING
CC   THE FORMATION OF THE NEUROMUSCULAR JUNCTION; STIMULATING
CC   LOBULOALVEOLAR BUDDING AND MILK PRODUCTION IN THE MAMMARY GLAND
CC   AND INDUCING DIFFERENTIATION OF MAMMARY TUMOR CELLS; STIMULATING
CC   SCHWANN CELL PROLIFERATION; IMPLICATION IN THE DEVELOPMENT OF THE
CC   MYOCARDIUM SUCH AS TRABECULATION OF THE DEVELOPING HEART.
CC   -!- SUBUNIT: THE CYTOPLASMIC DOMAIN INTERACTS WITH THE LIM DOMAIN
CC   REGION OF LIMK1 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
CC   A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
CC   MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE. THE SECRETED
CC   ISOFORM GGF2 HAS A SIGNAL PEPTIDE. THE ISOFORM BETA3 MAY BE
CC   NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 10 ISOFORMS; ALPHA (SHOWN HERE);
CC   ALPHA1A, ALPHA2B, ALPHA3, BETA1, BETA1A, BETA1A, BETA2, BETA3/GGFHFB1,

CC   -!- GGF2/GGFHPP2 AND SMDF (AC Q15491); ARE PRODUCED BY ALTERNATIVE
CC   SPLICING. THEY HAVE BEEN CLASSIFIED AS TYPE I NRGS (VARIANTS WITH
CC   AN IG DOMAIN AND A GLYCOSYLATION DOMAIN; ALPHA AND BETA), TYPE II
CC   NRGS (VARIANTS WITH AN IG DOMAIN BUT NO GLYCOSYLATION DOMAIN;
CC   GGF2) AND TYPE III NRGS (VARIANTS WITH A CYS-RICH DOMAIN; SMDF).
CC   ALL THESE ISOFORMS PERFORM DISTINCT TISSUE-SPECIFIC FUNCTIONS.
CC   -!- TISSUE SPECIFICITY: TYPE I ISOFORMS ARE THE PREDOMINANT FORMS
CC   EXPRESSED IN THE ENDOCARDIUM. ISOFORM ALPHA IS EXPRESSED IN
CC   BREAST, OVARY, TESTIS, PROSTATE, HEART, SKELETAL MUSCLE, LUNG,
CC   PLACENTA LIVER, KIDNEY, SALIVARY GLAND, SMALL INTESTINE AND BRAIN,
CC   BUT NOT IN UTERUS, STOMACH, PANCREAS, AND SPLEEN. ALPHA2 IS THE
CC   PREDOMINANT FORM IN MESENCHYMAL CELLS AND IN NONNEURONAL ORGANS,
CC   WHEREAS BETA1 IS THE MAJOR NEURONAL FORM. BETA3 IS EXPRESSED IN
CC   SPINAL CORD AND BRAIN. GGF2 IS THE MAJOR FORM IN SKELETAL MUSCLE
CC   CELLS; IN THE NERVOUS SYSTEM IT IS EXPRESSED IN SPINAL CORD AND
CC   BRAIN. ALSO DETECTED IN ADULT HEART, PLACENTA, LUNG, LIVER,
CC   KIDNEY, AND PANCREAS.
CC   -!- DEVELOPMENTAL STAGE: DETECTABLE AT EARLY EMBRYONIC AGES.
CC   -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC   OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC   PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC   DIMERIZATION (BY SIMILARITY).
CC   -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC   DOMAIN.
CC   -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
CC   SIMILARITY).
CC   -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
CC   EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC   FORM.
CC   -!- DISEASE: INVOLVED IN A RARE T(8;11) CHROMOSOMAL TRANSLOCATION THAT
CC   FUSES THE 5'END OF DDZ4 TO NRG1 (ISOFORM BETA3). THE PRODUCT OF
CC   THIS TRANSLOCATION HAS FIRST BE THOUGHT TO BE AN ALTERNATIVE
CC   SPLICED ISOFORM, CALLED GAMMA-HEREGULIN. GAMMA-HEREGULIN IS A
CC   SOLUBLE ACTIVATING LIGAND FOR THE ERBB2-ERBB3 RECEPTOR COMPLEX AND
CC   ACTS AS AN AUTOCRINE GROWTH FACTOR IN A SPECIFIC BREAST CANCER
CC   CELL LINE (MDA-MB-175). NOT DETECTED IN BREAST CARCINOMA SAMPLES,
CC   INCLUDING DUCTAL, LOBULAR, MEDULLARY, AND MUCINOUS HISTOLOGICAL
CC   TYPES, NEITHER IN OTHER BREAST CANCER CELL LINES.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M94165; AAA58638.1; -.
DR   EMBL; M94166; AAA58639.1; -.
DR   EMBL; M94167; AAA58640.1; -.
DR   EMBL; M94168; AAA58641.1; -.
DR   EMBL; L12261; AAB59358.1; -.
DR   EMBL; U02325; AAA19950.1; -.
DR   EMBL; U02326; AAA19951.1; -.
DR   EMBL; U02327; AAA19952.1; -.
DR   EMBL; U02328; AAA19953.1; -.
DR   EMBL; U02329; AAA19954.1; -.
DR   EMBL; U02330; AAA19955.1; -.
DR   EMBL; L12260; AAB59622.1; -.
DR   EMBL; AF026146; AAD01795.1; -.
DR   EMBL; AF009227; AAC51756.1; ALT_INIT.
DR   PDB; 1HRE; 15-OCT-94.
DR   PDB; 1HRF; 15-OCT-94.
DR   MIM; 142445; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR002154; Neuregulin.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00047; Ig; 1.
DR   Pfam; PF02158; Neuregulin; 1.

```
DR   PRINTS; PR01089; NEUREGULIN.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW   Transmembrane; Multigene family; Polymorphism; 3D-structure;
KW   Alternative splicing; Chromosomal translocation.
FT   INIT_MET

Qy     1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQ  45
Db   177 HLVKCAEEKKTFCVNGGECFMVKDLSNPSRYLCKCQPGFTGARCTE 222

Query Match       41.0%;  Score 113.5;  DB 1;  Length 639;
Best Local Similarity  34.8%;  Pred. No. 2.2e-06;
Matches 16;  Conservative 14;  Mismatches 15;  Indels 1;  Gaps 1;
```

```
RESULT 5
NRG4_MOUSE
ID   NRG4_MOUSE          STANDARD;      PRT;   115 AA.
AC   Q9WTX4;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PRO-NEUREGULIN-4, SHORT ISOFORM (PRO-NRG4) (CONTAINS: NEUREGULIN-4
DE   (NRG-4)].
GN   NRG4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Liver;
RX   MEDLINE=99276098; PubMed=10348342;
RA   Harari D., Tzahar E., Romano J., Shelly M., Pierce J.H., Andrews G.C.,
RA   Yarden Y.;
RT   "Neuregulin-4: a novel growth factor that acts through the ErbB-4
RT   receptor tyrosine kinase.";
RL   Oncogene 18:2681-2689(1999).
CC   -!- FUNCTION: LOW AFFINITY LIGAND FOR THE ERBB4 TYROSINE KINASE
CC       RECEPTOR. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
CC       RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC       ACTIVATION OF THE ERBB RECEPTORS. DOES NOT BIND TO THE ERBB1,
CC       ERBB2 AND ERBB3 RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
CC       A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
CC       MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS MAY BE PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PANCREAS; WEAKLY EXPRESSED
CC       IN MUSCLE.
CC   -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC       OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC       PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC       DIMERIZATION (BY SIMILARITY).
CC   -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC       DOMAIN (BY SIMILARITY).
CC   -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
CC       EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC       FORM (BY SIMILARITY).
CC   -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
```

```
DR   EMBL; AF083067; AAD21874.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001336; EGF_1.
DR   Pfam; PF00008; EGF; 1.
DR   PRINTS; PR00009; EGFTGF.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
KW   Growth factor; EGF-like domain; Glycoprotein; Transmembrane;
KW   Multigene family; Alternative splicing.
FT   CHAIN         1    115   PRO-NEUREGULIN-4, MEMBRANE-BOUND FORM.
FT   CHAIN         1     61   NEUREGULIN-4.
FT   DOMAIN       62    115   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     63     83   INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT   DOMAIN       84    115   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        5     46   EGF-LIKE.
FT   DISULFID      9     23   BY SIMILARITY.
FT   DISULFID     17     34   BY SIMILARITY.
FT   DISULFID     36     45   BY SIMILARITY.
FT   CARBOHYD     39     39   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     60     60   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   115 AA;  12743 MW;  989A1E376F857B49 CRC64;
```

```
Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHHCRCKEGYQGVRCDQ 45
Db   5 HEQPCGPRHRSFCLNGGICYVIPTIPS--PFCRCIENYTGARCEE  47

Query Match       39.7%;  Score 110;  DB 1;  Length 115;
Best Local Similarity  42.2%;  Pred. No. 1.2e-06;
Matches 19;  Conservative 8;  Mismatches 16;  Indels 2;  Gaps 1;
```

```
RESULT 6
SMDF_HUMAN
ID   SMDF_HUMAN          STANDARD;      PRT;   296 AA.
AC   Q15491;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NEUREGULIN-1, SENSORY AND MOTOR NEURON-DERIVED FACTOR ISOFORM.
GN   NRG1 OR HGL OR NDF OR HRGA OR GGF OR SMDF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain stem, and Cerebellum;
RX   MEDLINE=95301541; PubMed=7782315;
RA   Ho W.-H., Armanini M.P., Nuijens A., Phillips H.S., Osheroff P.L.;
RT   "Sensory and motor neuron-derived factor. A novel heregulin variant
RT   highly expressed in sensory and motor neurons.";
RL   J. Biol. Chem. 270:14523-14532(1995).
CC   -!- FUNCTION: THE ISOFORM SMDF MAY PLAY A ROLE IN MOTOR AND SENSORY
CC       NEURON DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: SECRETED. MAY POSSESS AN INTERNAL UNCLEAVED
CC       SIGNAL SEQUENCE.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 10 ISOFORMS OF NRG1 ARE PRODUCED BY
CC       ALTERNATIVE SPLICING. EXCEPT FOR SMDF THEY ARE IN ENTRY AC Q02297.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN NERVOUS SYSTEM: SPINAL CORD MOTOR
CC       NEURONS, DORSAL ROOT GANGLION NEURONS, AND BRAIN. PREDOMINANT
CC       ISOFORM EXPRESSED IN SENSORY AND MOTOR NEURONS. NOT DETECTED IN
CC       ADULT HEART, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, AND
CC       PANCREAS. NOT EXPRESSED IN FETAL LUNG, LIVER, AND KIDNEY.
CC   -!- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED IN DEVELOPING SPINAL MOTOR
CC       NEURONS AND IN DEVELOPING CRANIAL NERVE NUCLEI. EXPRESSION IS
CC       MAINTAINED ONLY IN BOTH ADULT MOTOR NEURONS AND DORSAL ROOT
CC       GANGLION NEURONS.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
```

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; L41827; AAC41764.1; -.
DR   MIM; 142445; -.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF_2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
KW   Growth factor; EGF-like domain; Transmembrane; Multigene family;
KW   Alternative splicing.
FT   TRANSMEM   76   100   INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT   DOMAIN   58   91   CYS-RICH.
FT   DOMAIN   211   232   SER/THR-RICH.
FT   DOMAIN   233   277   EGF-LIKE.
FT   DISULFID   237   251   BY SIMILARITY.
FT   DISULFID   245   265   BY SIMILARITY.
FT   DISULFID   267   276   BY SIMILARITY.
SQ   SEQUENCE   296 AA; 31685 MW; 8D4174321F7EB02 CRC64;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL 47
             ||||||||||||||||||| ||||||||||| ||||||||||||||| ::
Db   233 HLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCKCPNEFTGDRCQNYV 280

Query Match   37.7%; Score 104.5; DB 1; Length 296;
Best Local Similarity   31.2%;
Matches   15; Conservative   14; Mismatches   18; Indels   1; Gaps   1;

RESULT 7
ID   NRG2_MOUSE   STANDARD;   PRT;   756 AA.
AC   P56974;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PRO-NEUREGULIN-2 PRECURSOR (PRO-NRG2) (CONTAINS: NEUREGULIN-2 (NRG-2).
DE   (DIVERGENT OF NEUREGULIN 1) (DON-1)].
GN   NRG2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS NRG2-5; NRG2-10 AND NRG2-16A).
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=97311398; PubMed=9160115;
RA   Carraway K.L. III, Weber J.L., Unger M.J., Ledesma J., Yu N.,
RA   Gassmann M., Lai C.;
RT   "Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine
RT   kinases.";
RL   Nature 387:512-516(1997).
RN   [2]
RP   SEQUENCE OF 150-756 FROM N.A. (ISOFORMS DON-1M AND DON-1S).
RC   TISSUE=Choroid plexus;
RX   MEDLINE=97342638; PubMed=9199335;
RA   Busfield S.J., Michnick D.A., Chickering T.W., Revett T.L., Ma J.,
RA   Woolf E.A., Comrack C.A., Dussault B.J., Woolf J., Goodearl A.D.J.,
RA   Gearing D.P.;
RT   "Characterization of a neuregulin-related gene, Don-1, that is highly
RT   expressed in restricted regions of the cerebellum and hippocampus.";
RL   Mol. Cell. Biol. 17:4007-4014(1997).
CC   -!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE
CC   RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
CC   RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC   ACTIVATION OF THE ERBB RECEPTORS. MAY ALSO PROMOTE THE

CC   -!- HETERODIMERIZATION WITH THE EGF RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
CC   A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
CC   MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS; DON-1M, DON-1S/NRG2-5,
CC   NRG2-10 AND NRG2-16A (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE
CC   SPLICING.
CC   -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN THE BRAIN, WITH LOWER
CC   LEVELS IN THE LUNG. IN THE CEREBELLUM, FOUND IN GRANULE AND
CC   PURKINJE CELLS.
CC   -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC   OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC   PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC   DIMERIZATION (BY SIMILARITY).
CC   -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC   DOMAIN (BY SIMILARITY).
CC   -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
CC   EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC   FORM (BY SIMILARITY).
CC   -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
CC   SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
DR   MGD; MGI:1098246; Nrg2.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR002154; Neuregulin.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00047; Ig; 1.
DR   Pfam; PF02158; Neuregulin; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW   Transmembrane; Multigene family; Alternative splicing.
FT   PROPEP   1   19
FT   CHAIN   20   756   PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.
FT   CHAIN   20   314   NEUREGULIN-2.
FT   DOMAIN   158   226   IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN   238   248   SER/THR-RICH.
FT   DOMAIN   249   290   EGF-LIKE.
FT   DOMAIN   627   633   POLY-PRO.
FT   TRANSMEM   316   336   BY SIMILARITY.
FT   DOMAIN   337   756   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN   20   19   EXTRACELLULAR (POTENTIAL).
FT   DISULFID   280   289   BY SIMILARITY.
FT   DISULFID   261   278   BY SIMILARITY.
FT   DISULFID   253   267   BY SIMILARITY.
FT   DISULFID   165   219   BY SIMILARITY.
FT   CARBOHYD   55   55   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   186   186   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   254   254   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   296   296   N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   280   280   C -> G (IN ISOFORM NRG2-10).
FT   VARSPLIC   281   756   MISSING (IN ISOFORM NRG2-10).
FT   VARSPLIC   282   330   VGYTGDRCQQFAMVNFSKHLGFELKEAEELYQKRVLTITGI
FT   VARSPLIC   331   756   CVALLVVG -> NGFFGQRCLEKLPLRLYMPDPKQSVLWDT
FT   VARSPLIC   331   756   PGTGVSSSQWSTSPSTLDLN (IN ISOFORM DON-1S).
FT   VARSPLIC   282   307   MISSING (IN ISOFORM DON-1S).
FT   VARSPLIC   282   307   VGYTGDRCQQFAMVNFSKHLGFELKE -> NGFFGQRCLEK
FT   VARSPLIC   LPLRLYMPDPKQK (IN ISOFORM DON-1M).
SQ   SEQUENCE   756 AA; 82213 MW; 51D85DC91BBE678E CRC64;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQF 46
             |:|| ||||||||||||| ||||||||||| ||| || || || ||
Db   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQF 46

Query Match   37.5%; Score 104; DB 1; Length 756;
Best Local Similarity   39.1%;
Matches   18; Conservative   8; Mismatches   18; Indels   2; Gaps   1;

```
Db      249  HARKCNEFAKSYCVNGGVCYYIEGI--NQLSCKCPVGYTGDRCQQF 292

RESULT   8

ID   NRG1_RAT       STANDARD;      PRT;    662 AA.
AC   P43322; P43323; P43324; P43325; P43326; P43327; P43328;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PRO-NEUREGULIN-1 PRECURSOR (PRO-NRG1) (CONTAINS: NEUREGULIN-1 (NEU
DE   DIFFERENTIATION FACTOR) (HEREGULIN) (HRG) (ACETYLCHOLINE RECEPTOR
DE   INDUCING ACTIVITY) (ARIA) (SENSORY AND MOTOR NEURON-DERIVED FACTOR)
DE   (GLIAL GROWTH FACTOR)].
GN   NRG1 OR NDF.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RC   TISSUE=Fibroblast;
RX   MEDLINE=94158863; PubMed=7509448;
RA   Wen D., Suggs S.V., Karunagaran D., Liu N., Cupples R.L., Luo Y.,
RA   Janssen A.M., Ben-Baruch N., Trollinger D.B., Jacobsen V.L.,
RA   Meng S.-Y., Lu H.S., Hu S., Chang D., Yang W., Yanigahara D.,
RA   Koski R.A., Yarden Y.;
RT   "Structural and functional aspects of the multiplicity of Neu
RT   differentiation factors.";
RL   Mol. Cell. Biol. 14:1909-1919(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA2C/NDF44), AND PARTIAL SEQUENCE.
RC   TISSUE=Fibroblast;
RX   MEDLINE=92257596; PubMed=1348215;
RA   Peles E., Cupples R., Suggs S.V., Bacus S.S., Koski R.A., Lu H.S.,
RA   Levy R.B., Yarden Y.;
RT   "Isolation of the neu/HER-2 stimulatory ligand: a 44 kd glycoprotein
RT   that induces differentiation of mammary tumor cells.";
RL   Cell 69:205-216(1992).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA2C/NDF44).
RC   TISSUE=Fibroblast;
RX   MEDLINE=92208945; PubMed=1349853;
RA   Wen D., Peles E., Cupples R., Suggs S.V., Bacus S.S., Luo Y.,
RA   Trail G., Hu S., Silbiger S.M., Levy R.B., Koski R.A., Lu H.S.,
RA   Yarden Y.;
RT   "Neu differentiation factor: a transmembrane glycoprotein containing
RT   an EGF domain and an immunoglobulin homology unit.";
RL   Cell 69:559-572(1992).
RN   [4]
RP   REGULATION OF PROCESSING (ISOFORM ALPHA2C/NDF44).
RX   MEDLINE=99069430; PubMed=9852099;
RA   Liu X., Hwang H., Cao L., Wen D., Liu N., Graham R.M., Zhou M.;
RT   "Release of the neuregulin functional polypeptide requires its
RT   cytoplasmic tail.";
RL   J. Biol. Chem. 273:34335-34340(1998).
RN   [5]
RP   INTERACTION WITH LIMK1.
RX   MEDLINE=98352096; PubMed=9685409;
RA   Wang J.Y., Frenzel K.E., Wen D., Falls D.L.;
RT   "Transmembrane neuregulins interact with LIM kinase 1, a cytoplasmic
RT   protein kinase implicated in development of visuospatial cognition.";
RL   J. Biol. Chem. 273:20525-20534(1998).
-!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE
-!-   RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
-!-   RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
-!-   ACTIVATION OF THE ERBB RECEPTORS. THE MULTIPLE ISOFORMS PERFORM
-!-   DIVERSE FUNCTIONS SUCH AS INDUCING GROWTH AND DIFFERENTIATION OF
-!-   EPITHELIAL, GLIAL, NEURONAL, AND SKELETAL MUSCLE CELLS; INDUCING
-!-   EXPRESSION OF ACETYLCHOLINE RECEPTOR IN SYNAPTIC VESICLES DURING
-!-   THE FORMATION OF THE NEUROMUSCULAR JUNCTION; STIMULATING
-!-   LOBULOALVEOLAR BUDDING AND MILK PRODUCTION IN THE MAMMARY GLAND
-!-   AND INDUCING DIFFERENTIATION OF MAMMARY TUMOR CELLS; STIMULATING
-!-   SCHWANN CELL PROLIFERATION; IMPLICATION IN THE DEVELOPMENT OF THE
-!-   MYOCARDIUM SUCH AS TRABECULATION OF THE DEVELOPING HEART (BY
-!-   SIMILARITY).
-!- SUBUNIT: THE CYTOPLASMIC DOMAIN INTERACTS WITH THE LIM DOMAIN
-!-   REGION OF LIMK1.
-!- SUBCELLULAR LOCATION: EXISTS AS A TYPE I MEMBRANE PROTEIN AND AS A
-!-   PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE MEMBRANE-
-!-   BOUND FORM DOES NOT SEEM TO BE ACTIVE.
-!- ALTERNATIVE PRODUCTS: AT LEAST 8 ISOFORMS; ALPHA2A/NDF38,
-!-   ALPHA2B/NDF19, ALPHA2C/NDF44, BETA1, BETA2/NDF40, BETA2A/NDF22,
-!-   BETA3/NDF4 AND BETA4/NDF42A (SHOWN HERE); ARE PRODUCED BY
-!-   ALTERNATIVE SPLICING.
-!- TISSUE SPECIFICITY: WIDELY EXPRESSED. MOST TISSUES CONTAIN ALPHA2A
-!-   AND ALPHA2B ISOFORMS. ALPHA2 AND BETA2 ARE THE PREDOMINANT FORMS
-!-   IN MESENCHYMAL AND NONNEURONAL ORGANS. BETA1 IS ENRICHED IN BRAIN
-!-   AND SPINAL CORD, BUT NOT IN MUSCLE AND HEART. ALPHA2C IS HIGHLY
-!-   EXPRESSED IN SPINAL CORD, MODERATELY IN LUNG, BRAIN, OVARY, AND
-!-   STOMACH, IN LOW AMOUNTS IN THE KIDNEY, SKIN AND HEART AND NOT
-!-   DETECTED IN THE LIVER, SPLEEN, AND PLACENTA.
-!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION OF
-!-   TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
-!-   PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
-!-   DIMERIZATION.
-!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
-!-   DOMAIN.
-!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
-!-   EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
-!-   FORM.
-!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE.
-!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
-!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
-!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
    ------------------------------------------------------------------------
    This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
    ------------------------------------------------------------------------
DR   EMBL; U02315; AAA19940.1; -.
DR   EMBL; U02316; AAA19941.1; -.
DR   EMBL; U02317; AAA19942.1; -.
DR   EMBL; U02318; AAA19943.1; -.
DR   EMBL; U02319; AAA19944.1; -.
DR   EMBL; U02320; AAA19945.1; -.
DR   EMBL; U02321; AAA19946.1; -.
DR   EMBL; U02322; AAA19947.1; -.
DR   EMBL; U02323; AAA19948.1; -.
DR   EMBL; U02324; AAA19949.1; -.
DR   EMBL; M92430; -; NOT_ANNOTATED_CDS.
DR   HSSP; Q02297; 1HRF.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001089; NEUREGULIN.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR002154; Neuregulin.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00047; Ig; 1.
DR   Pfam; PF02158; Neuregulin; 1.
DR   PRINTS; PR01089; NEUREGULIN.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00408; IgC2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
KW   Growth factor; Immunoglobulin domain; Glycoprotein;
KW   Transmembrane; Multigene family; Alternative splicing.
FT   PROPEP       1     13        PRO-NEUREGULIN-1.
FT   CHAIN       14    662        NEUREGULIN-1, MEMBRANE-BOUND FORM.
FT   CHAIN       14    264        NEUREGULIN-1.
FT   DOMAIN     264    265        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   266    288        INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT   DOMAIN     289    662        CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      50    119        IG-LIKE C2-TYPE DOMAIN.
```

```
RESULT 9
NRG1_CHICK

ID   NRG1_CHICK     STANDARD;      PRT;   602 AA.
AC   O05199; O73750; O73751; O73752;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PRO-NEUREGULIN-1 PRECURSOR (PRO-NRG1) [CONTAINS: NEUREGULIN-1
DE   (ACETYLCHOLINE RECEPTOR INDUCING ACTIVITY) (ARIA)].
GN   NRG1 OR ARIA.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS ARIA), AND PARTIAL SEQUENCE.
RC   STRAIN=WHITE LEGHORN; TISSUE=Brain;
RX   MEDLINE=93201602; PubMed=8453670;
RA   Falls D.L., Rosen K.M., Corfas G., Lane W.S., Fischbach G.D.;
RT   "ARIA, a protein that stimulates acetylcholine receptor synthesis, is
RT   a member of the neu ligand family.";
RL   Cell 72:801-815(1993).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS BETA1A, BETA2A AND BETA2B).
RC   TISSUE=Brain, and Spinal cord;
RX   MEDLINE=98150951; PubMed=9491987;
RA   Yang X., Kuo Y., Devay P., Yu C., Role L.;
RT   "A cysteine-rich isoform of neuregulin controls the level of
RT   expression of neuronal nicotinic receptor channels during
RT   synaptogenesis.";
RL   Neuron 20:255-270(1998).
```

```
Qy   1    HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL  47
          |  ||  :| :|  |||| ||||  :|    ||   || | || | ||  :
Db   178  HLIKCAEKEKTFCVNGGECFTVKDLSNPSRYLCKCPNEFTGDRCQNYV  225
```

```
Query Match     37.0%;  Score 102.5;  DB 1;  Length 662;
Best Local Similarity   31.0%;
Matches   15;  Conservative  13;  Mismatches  19;  Indels  1;  Gaps  1;
                 Pred. No. 4.7e-05;
```

```
CC   -!- FUNCTION: DIRECT LIGAND FOR THE ERBB TYROSINE KINASE RECEPTORS.
CC       THE MULTIPLE ISOFORMS PERFORM DIVERSE FUNCTIONS: CYSTEIN-RICH
CC       DOMAIN CONTAINING ISOFORMS (CRD-NRG) PROBABLY REGULATE THE
CC       EXPRESSION OF NICOTINIC ACETYLCHOLINE RECEPTORS AT DEVELOPING
CC       INTERNEURONAL SYNAPSES. THE IG-NRG ISOFORM IS REQUIRED FOR THE
CC       INITIAL INDUCTION AND/OR MAINTENANCE OF THE MATURE LEVELS OF
CC       ACETYLCHOLINE RECEPTORS AT NEUROMUSCULAR SYNAPSES.
CC   -!- SUBCELLULAR LOCATION: EXISTS AS A TYPE I MEMBRANE PROTEIN AND AS A
CC       PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE MEMBRANE-
CC       BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS; ARIA/IG-NRG (SHOWN
CC       HERE), CRD-NRG-BETA1A, CRD-NRG-BETA2A AND CRD-NRG-BETA2B; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING. ARIA CONTAINS AN IG-LIKE DOMAIN,
CC       WHEREAS IN THE CRD-NRG (OR NARIA) ISOFORMS, THE EGF-LIKE DOMAIN IS
CC       REPLACED BY A CYSTEINE-RICH DOMAIN (CRD).
CC   -!- DEVELOPMENTAL STAGE: CRD-NRG ISOFORM IS DETECTED AT EMBRYONIC DAY
CC       4 (ED4) IN BOTH VISCERAL AND SOMATIC MOTOR NEURONS OF SPINAL CORD
CC       AND IS HIGHEST AT ED6. IG-NRG ISOFORM IS NOT EXPRESSED UNTIL ED 6
CC       IN SPINAL CORD. AT ED 11 BOTH ISOFORMS DISPLAY COMPARABLE LEVELS.
CC   -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC       OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC       PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC       DIMERIZATION (BY SIMILARITY).
CC   -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC       DOMAIN.
CC   -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
CC       EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC       FORM.
CC   -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
```

```
DR   EMBL; L11264; AAA49037.1; -.
DR   EMBL; AF045654; AAC05670.1; -.
DR   EMBL; AF045655; AAC05671.1; -.
DR   EMBL; AF045656; AAC05672.1; -.
DR   HSSP; Q02297; 1HRF.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000306; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR002154; Neuregulin.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF02158; Neuregulin; 1.
DR   PRINTS; PR01089; NEUREGULIN.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
KW   Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW   Transmembrane; Alternative splicing.
FT   CHAIN         1    602       PRO-NEUREGULIN-1.
FT   CHAIN         1    205       NEUREGULIN-1.
FT   DOMAIN        1    205       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    207    229       INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT   DOMAIN      230    602       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       42    112       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      125    136       SER/THR-RICH.
FT   DOMAIN      137    181       EGF-LIKE.
FT   DISULFID     49    105       BY SIMILARITY.
FT   DISULFID    141    155       BY SIMILARITY.
FT   DISULFID    149    169       BY SIMILARITY.
FT   DISULFID    171    180       BY SIMILARITY.
FT   CARBOHYD     21     21       N-LINKED (GLCNAC...) (POTENTIAL).
```

```
FT   DOMAIN      165    177       SER/THR-RICH.
FT   DOMAIN      178    222       EGF-LIKE.
FT   DISULFID     57    112
FT   DISULFID    182    196       BY SIMILARITY.
FT   DISULFID    190    210       BY SIMILARITY.
FT   DISULFID    212    221       BY SIMILARITY.
FT   CARBOHYD    120    120       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    126    126       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    164    164       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC    213    256       PNEFTGDRCQNYVMASFYMTSRRKRQETEKPLERKLDHSLIV
FT                                KES -> QPGFTGARCTENVPMKVQTQE (IN ISOFORM
FT                                ALPHA2A, ISOFORM ALPHA2B AND ISOFORM
FT                                ALPHA2C).
FT   VARSPLIC    231    257       MTSRRKRQETEKPLERKLDHSLVVKESK -> KHLGIEFME
FT                                (IN ISOFORM BETA1).
FT   VARSPLIC    231    241       MTSRRKRQETE -> STSTPFLSLPE (IN ISOFORM
FT                                BETA3).
FT   VARSPLIC    242    662       MISSING (IN ISOFORM BETA3).
FT   VARSPLIC    231    256       MISSING (IN ISOFORM BETA2 AND ISOFORM
FT                                BETA2A).
FT   VARSPLIC    325    330       PPENVQ -> RVRTRG (IN ISOFORM BETA2).
FT   VARSPLIC    446    662       MISSING (IN ISOFORM ALPHA2C).
FT   VARSPLIC    446    484       YVSAMTTPARMSPVDFHTPSSPRKSPPSEMSPPVSSMTVS
FT                                -> HNLIAELRRNKAYRSKCMQIOLSATHLRPSSITHLGFI
FT                                L (IN ISOFORM ALPHA2B).
FT   CONFLICT    485    662       MISSING (IN ISOFORM ALPHA2B).
FT   CONFLICT     90     90       K -> N (IN REF. 2).
FT   CONFLICT    137    137       T -> I (IN REF. 2).
FT   CONFLICT    208    208       Y -> S (IN REF. 2; AA SEQUENCE).
SQ   SEQUENCE   662 AA;  73288 MW;  1C31ABCF2A8EB1D5 CRC64;
```

```
FT   CARBOHYD    113    113   N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD    126    126   N-LINKED (GLCNAC...)    (POTENTIAL).
FT   VARSPLIC      1    127   MWATSEGPLQYSLAPTQYDVNSSYNTVPPKLKEMKNQEVAV
FT                           GQKVLRCETTSEYPALRFKWLKNGKEITKKNRPENVKIPK
FT                           KQRKYSELHIYRATALADAGEYACRVSSKLGNDSTKASVIIT
FT                           DTNA -> MSEVGTETPPSPSAQLSPDASLGGLPAEENMPG
FT                           PHREDSRVPGVAGLASTCCVCLEAERLRGCLNSEKICIAPI
FT                           LACLLSLCLIACGLKWFVDKIPEYDSPTHLDPGRIGQDDP
FT                           STVDPTALSAWPSEVVASPFPIPSLESKAEYTVQTDSSLV
FT                           PSRPFLQPSLYNRILDVGLMKSSATPSLSPSSLEPTTASQAO
FT                           ATETNLQTAPKLS (IN ISOFORM BETA1A, ISOFORM
FT                           BETA2A AND ISOFORM BETA2B).
FT   VARSPLIC    191    198   MISSING (IN ISOFORM BETA2A AND ISOFORM
FT                           BETA2B).
FT   VARSPLIC    388    405   VSAMTTPARMSPVDFHTP -> HTPPTSLLLAGKVSLRVS
FT                           (IN ISOFORM BETA2B).
FT   VARSPLIC    406    602   MISSING (IN ISOFORM BETA2B).
SQ   SEQUENCE   602 AA;  67453 MW;  4183C0E56GCE5D346 CRC64;
```

```
Qy   Query Match   35.6%;  Score 98.5;  DB 1;  Length 602;
     Best Local Similarity  33.3%;  Pred. No. 0.00013;
     Matches  16;  Conservative  11;  Mismatches  20;  Indels  1;  Gaps  1;

Qy     1  HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL   47
Db   137  HLTKCDIKQKAPCVNGGECYMVKDLPNPPRYLCRCRNEFTGDRCQNYV  184
```

```
RESULT 10
NRG2_HUMAN
ID   NRG2_HUMAN     STANDARD;      PRT;   850 AA.
AC   O14511;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PRO-NEUREGULIN-2 PRECURSOR (PRO-NRG2) [CONTAINS: NEUREGULIN-2 (NRG-2)
DE   (NEURAL-AND THYMUS-DERIVED ACTIVATOR FOR ERBB KINASES) (NTAK)
DE   (DIVERGENT OF NEUREGULIN 1) (DON-1)].
GN   NRG2 OR NTAK.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Neuroblastoma;
RX   MEDLINE=98006324; PubMed=9348101;
RA   Higashiyama S., Horikawa M., Yamada K., Ichino N., Nakano N.,
RA   Nakagawa T., Miyagawa J., Matsushita N., Nagatsu T., Taniguchi N.,
RA   Ishiguro H.;
RT   "A novel brain-derived member of the epidermal growth factor family
RT   that interacts with ErbB3 and ErbB4.";
RL   J. Biochem. 122:675-680(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS DON-1B AND DON-1R).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97342638; PubMed=9199335;
RA   Busfield S.J., Michnick D.A., Chickering T.W., Revett T.L., Ma J.,
RA   Woolf E.A., Comrack C.A., Dussault B.J., Woolf J., Goodearl A.D.J.,
RA   Gearing D.P.;
RT   "Characterization of a neuregulin-related gene, Don-1, that is highly
RT   expressed in restricted regions of the cerebellum and hippocampus.";
RL   Mol. Cell. Biol. 17:4007-4014(1997).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4; 5 AND 6).
RC   TISSUE=Lung, and Fetal brain;
RX   MEDLINE=99295836; PubMed=10369162;
RA   Ring H.Z., Chang H., Guilbot A., Brice A., LeGuern E., Francke U.;
RT   "The human neuregulin 2 (NRG2) gene: cloning, mapping and evaluation
RT   as a candidate for the autosomal recessive form of Charcot-Marie-Tooth
RT   disease linked to 5q.";
RL   Hum. Genet. 104:326-332(1999).
```

```
CC   -!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE
CC       RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
CC       RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC       ACTIVATION OF THE ERBB RECEPTORS. MAY ALSO PROMOTE THE
CC       HETERODIMERIZATION WITH THE EGF RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
CC       A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
CC       MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 8 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4, 5, 6,
CC       DON-1B AND DON-1R; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: RESTRICTED TO THE CEREBELLUM IN THE ADULT.
CC   -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC       OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC       PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC       DIMERIZATION (BY SIMILARITY).
CC   -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC       DOMAIN (BY SIMILARITY).
CC   -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
CC       EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC       FORM (BY SIMILARITY).
CC   -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; AB005060; BAA23417.1; -.
DR   EMBL; AF119162; AAF28848.1; -.
DR   EMBL; AF119151; AAF28848.1; JOINED.
DR   EMBL; AF119152; AAF28848.1; JOINED.
DR   EMBL; AF119153; AAF28848.1; JOINED.
DR   EMBL; AF119154; AAF28848.1; JOINED.
DR   EMBL; AF119155; AAF28848.1; JOINED.
DR   EMBL; AF119156; AAF28848.1; JOINED.
DR   EMBL; AF119157; AAF28848.1; JOINED.
DR   EMBL; AF119158; AAF28848.1; JOINED.
DR   EMBL; AF119159; AAF28848.1; JOINED.
DR   EMBL; AF119160; AAF28848.1; JOINED.
DR   EMBL; AF119161; AAF28848.1; JOINED.
DR   EMBL; AF119162; AAF28849.1; -.
DR   EMBL; AF119151; AAF28849.1; JOINED.
DR   EMBL; AF119152; AAF28849.1; JOINED.
DR   EMBL; AF119153; AAF28849.1; JOINED.
DR   EMBL; AF119154; AAF28849.1; JOINED.
DR   EMBL; AF119155; AAF28849.1; JOINED.
DR   EMBL; AF119156; AAF28849.1; JOINED.
DR   EMBL; AF119157; AAF28849.1; JOINED.
DR   EMBL; AF119158; AAF28849.1; JOINED.
DR   EMBL; AF119159; AAF28849.1; JOINED.
DR   EMBL; AF119160; AAF28849.1; JOINED.
DR   EMBL; AF119161; AAF28849.1; JOINED.
DR   EMBL; AF119162; AAF28850.1; -.
DR   EMBL; AF119151; AAF28850.1; JOINED.
DR   EMBL; AF119152; AAF28850.1; JOINED.
DR   EMBL; AF119153; AAF28850.1; JOINED.
DR   EMBL; AF119154; AAF28850.1; JOINED.
DR   EMBL; AF119155; AAF28850.1; JOINED.
DR   EMBL; AF119156; AAF28850.1; JOINED.
DR   EMBL; AF119157; AAF28850.1; JOINED.
DR   EMBL; AF119158; AAF28850.1; JOINED.
DR   EMBL; AF119159; AAF28850.1; JOINED.
DR   EMBL; AF119160; AAF28850.1; JOINED.
DR   EMBL; AF119161; AAF28850.1; JOINED.
DR   EMBL; AF119162; AAF28851.1; -.
DR   EMBL; AF119151; AAF28851.1; JOINED.
DR   EMBL; AF119152; AAF28851.1; JOINED.
DR   EMBL; AF119153; AAF28851.1; JOINED.
DR   EMBL; AF119154; AAF28851.1; JOINED.
DR   EMBL; AF119155; AAF28851.1; JOINED.
DR   EMBL; AF119156; AAF28851.1; JOINED.
DR   EMBL; AF119157; AAF28851.1; JOINED.
```

```
DR   EMBL; AF119158; AAF28851.1; JOINED.
DR   EMBL; AF119159; AAF28851.1; JOINED.
DR   EMBL; AF119160; AAF28851.1; JOINED.
DR   EMBL; AF119161; AAF28851.1; JOINED.
DR   EMBL; AF119158; AAF28852.1; JOINED.
DR   EMBL; AF119151; AAF28852.1; JOINED.
DR   EMBL; AF119152; AAF28852.1; JOINED.
DR   EMBL; AF119153; AAF28852.1; JOINED.
DR   EMBL; AF119154; AAF28852.1; JOINED.
DR   EMBL; AF119155; AAF28852.1; JOINED.
DR   EMBL; AF119156; AAF28852.1; JOINED.
DR   EMBL; AF119157; AAF28853.1; JOINED.
DR   EMBL; AF119151; AAF28853.1; JOINED.
DR   EMBL; AF119152; AAF28853.1; JOINED.
DR   EMBL; AF119153; AAF28853.1; JOINED.
DR   EMBL; AF119154; AAF28853.1; JOINED.
DR   EMBL; AF119155; AAF28853.1; JOINED.
DR   EMBL; AF119156; AAF28853.1; JOINED.
DR   MIM; 603818; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR002154; Neuregulin.
DR   Pfam; PF02158; Neuregulin; 2.
DR   Pfam; PF00047; Ig; 1.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
KW   Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW   Transmembrane; Multigene family; Alternative splicing.
FT   PROPEP         1    111
FT   CHAIN        112    850       NEUREGULIN-2.
FT   CHAIN        112    404       PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.
FT   CHAIN        112    405       NEUREGULIN-2.
FT   DOMAIN       112    406       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     406    426       INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT   DOMAIN       427    850       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       250    318       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN       330    340       SER/THR-RICH.
FT   DOMAIN       341    382       EGF-LIKE.
FT   DOMAIN        10     13       POLY-PRO.
FT   DOMAIN        20     30       POLY-PRO.
FT   DOMAIN        33     47       POLY-SER.
FT   DOMAIN        87     90       POLY-ALA.
FT   DOMAIN       721    727       POLY-PRO.
FT   DISULFID     257    311       BY SIMILARITY.
FT   DISULFID     345    359       BY SIMILARITY.
FT   DISULFID     353    370       BY SIMILARITY.
FT   DISULFID     372    381       BY SIMILARITY.
FT   CARBOHYD      52     52       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      53     53       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     147    147       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     278    278       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     346    346       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC       1    233       MROVCCSALPPPLEKGRCSSYSDSSSSSSERSSSSSSSSSS
FT                                 ESGSSSRSSSNNSSISRPAPPEPSMLLFGVSLIACYSPSLKSVQDQ
FT                                 ARSRAAAAGGMWRDPAPGFSMLLFGVSLIACYSPSLKSVQDQ
FT                                 AYKAPVVVEGKRVQGLVPAGSSSVPLERNQRYIFFLEPTEOPL
FT                                 DKWPLRSGGLQREQVTSVGSGSCVPLERNQRYIFFLEPTEOPL
FT                                 VFKTAPALDTNGKNLKKEVGKILCTDC -> MSESRRGR
FT                                 GRGKHPEGRKREREPDPGEK (IN ISOFORM
FT                                 1R).
FT   VARSPLIC       1    241       MISSING (IN ISOFORM DON-1B).
FT   VARSPLIC     374    396       NGFPGQRCLEKLPLRIYMPDPKQ -> VGTTGDRCQQFAMV
FT                                 NFS (IN ISOFORM 2).
FT   VARSPLIC     374    397       NGFPGQRCLEKLPLRIYMPDPKQK -> VGYTGDRCQQFAM
FT                                 VNFSKHLGFELKE (IN ISOFORM 4).
FT   VARSPLIC     397    397       K -> KHLGFELKE (IN ISOFORM 3 AND IN
FT                                 ISOFORM DON-1B).
FT   VARSPLIC     397    422       KAEELYQKRVLTITGICVALLVVGIV -> SVLMDTPGTGV
FT                                 SSSQWSTSPSTLDLN (IN ISOFORM 6).
FT   VARSPLIC     423    850       MISSING (IN ISOFORM 6).
FT   VARSPLIC     397    426       KAEELYQKRVLTITGICVALLVVGIVCVVA -> SVLMDTP
FT   VARSPLIC     427    850       GTGVSSSQWSTSPKPRSCTRRGS (IN ISOFORM 5).
FT   VARSPLIC     427    850       MISSING (IN ISOFORM 5).
SQ   SEQUENCE   850 AA;  91678 MW;  7124C089435FD2P4 CRC64;
```

```
Query Match             32.5%;  Score 90;  DB 1;  Length 850;
Best Local Similarity   34.9%;  Pred. No. 0.0019;
Matches   15;  Conservative    9;  Mismatches   17;  Indels    2;  Gaps    1;

Qy      1 HFKPCRDKDLAYCLNDGECFVIETLLRGSHKHCRCKEGYQGVRC 43
          |.|.|.|..|||.||||..|.|..||::.||.||:|.:|:||||||
Db    341 HARKCNETAKSYCVNGGVCYYIEGI--NQLSCKCPNGFFGQRC 381
```

```
RESULT 11
NRG2_RAT
ID   NRG2_RAT      STANDARD;      PRT;    868 AA.
AC   O35569; O35570; O35571; O35572; O35073;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PRO-NEUREGULIN-2 PRECURSOR (PRO-NRG2) (CONTAINS: NEUREGULIN-2 (NRG-2)
DE   (NEURAL-AND THYMUS-DERIVED ACTIVATOR FOR ERBB KINASES) (NTAK)).
GN   NRG2 OR NTAK.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 128-162, AND ALTERNATIVE SPLICING.
RX   MEDLINE=98006324; PubMed=9348101;
RA   Higashiyama S., Horikawa M., Yamada K., Ichino N., Nakano N.,
RA   Nakagawa T., Miyagawa J., Matsushita N., Nagatsu T., Taniguchi N.,
RA   Ishiguro H.;
RT   "A novel brain-derived member of the epidermal growth factor family
RT   that interacts with ErbB3 and ErbB4.";
RL   J. Biochem. 122:675-680(1997).
RN   [2]
RP   SEQUENCE OF 109-868 FROM N.A. (ISOFORMS NRG2-ALPHA AND NRG2-BETA).
RC   TISSUE=Cerebellum;
RX   MEDLINE=97311397; PubMed=9168114;
RA   Chang H., Riese D.J. II, Gilbert W., Stern D.F., McMahan U.J.;
RT   "Ligands for ErbB-family receptors encoded by a neuregulin-like
RT   gene.";
RL   Nature 387:509-512(1997).
CC   -!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE
CC       RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
CC       RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC       ACTIVATION OF THE ERBB RECEPTORS. MAY ALSO PROMOTE THE
CC       HETERODIMERIZATION WITH THE EGF RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
CC       A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
CC       MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS; NTAK-ALPHA1 (SHOWN
CC       HERE), NTAK-ALPHA2A, NTAK-ALPHA2B/NTAK-ALPHA2-1P, NTAK-BETA, NTAK-
CC       GAMMA, NRG2-ALPHA AND NRG2-BETA; ARE PRODUCED BY ALTERNATIVE
CC       SPLICING. THE ALPHA-TYPE AND BETA-TYPE DIFFER IN THE EGF-LIKE
CC       DOMAIN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN MOST PARTS OF THE BRAIN,
CC       ESPECIALLY THE OLFACTORY BULB AND CEREBELLUM WHERE IT LOCALIZES IN
CC       GRANULE AND PURKINJE CELLS. IN THE HIPPOCAMPUS, FOUND IN THE
CC       GRANULE CELLS OF THE DENTRATE GYRUS. IN THE BASAL FOREBRAIN, FOUND
CC       IN THE CHOLINERGIC CELLS. IN THE HINDBRAIN, WEAKLY DETECTABLE IN
CC       THE MOTOR TRIGEMINAL NUCLEUS. NOT DETECTED IN THE HYPOTHALAMUS.
CC       ALSO FOUND IN THE LIVER AND IN THE THYMUS. NOT DETECTED IN HEART,
CC       ADRENAL GLAND, OR TESTIS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE EMBRYO, EXPRESSED IN THE BRAIN OF
CC       E11.5 EMBRYOS WHERE IT IS FOUND IN THE TELENCEPHALON, BUT NOT IN
CC       THE HINDBRAIN. NOT FOUND IN THE HEART. IN THE ADULT, FOUND IN
CC       BRAIN AND THYMUS.
CC   -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC       OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
```

```
CC       PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC       DIMERIZATION (BY SIMILARITY).
-!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
         DOMAIN (BY SIMILARITY).
-!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
         EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
         FORM (BY SIMILARITY).
-!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
         SIMILARITY).
-!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
-!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
-!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
---------------------------------------------------------------------------
DR   EMBL; D89995; BAA23344.1; -.
DR   EMBL; D89996; BAA23345.1; -.
DR   EMBL; D89997; BAA23346.1; -.
DR   EMBL; D89998; BAA23347.1; -.
DR   EMBL; AB001576; BAA23348.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR002154; Neuregulin.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00047; Ig; 1.
DR   Pfam; PF02158; Neuregulin; 2.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW   Transmembrane; Multigene family; Alternative splicing.
FT   PROPEP        1    127
FT   CHAIN       128    868
FT   CHAIN       128    428
FT   CHAIN       128    429
FT   DOMAIN      128    450       NEUREGULIN-2.
FT   TRANSMEM    430    450       INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT   DOMAIN      451    868       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      266    334       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      346    356       SER/THR-RICH.
FT   DOMAIN      357    398       EGF-LIKE.
FT   DOMAIN       22     32       POLY-SER.
FT   DOMAIN       35     45       POLY-SER.
FT   DOMAIN       56     59       POLY-THR.
FT   DOMAIN      103    106       POLY-ALA.
FT   DOMAIN      739    745       POLY-PRO.
FT   DISULFID    273    327       BY SIMILARITY.
FT   DISULFID    361    375       BY SIMILARITY.
FT   DISULFID    369    386       BY SIMILARITY.
FT   DISULFID    388    397       BY SIMILARITY.
FT   CARBOHYD     33     33       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     34     34       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    163    163       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    294    294       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    362    362       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      1      1       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC    108    108       MISSING (IN ISOFORM NRG2-BETA).
FT   VARSPLIC    220    222       PLV -> FFF (IN ISOFORM NTAK-ALPHA2-1P).
FT   VARSPLIC    388    388       C -> G (IN ISOFORM NTAK-GAMMA).
FT   VARSPLIC    389    868       MISSING (IN ISOFORM NTAK-GAMMA).
FT   VARSPLIC    390    412       NGFFGQRCLEKLPLRLYMPDPKQ -> VGYTGDRCQQFAMV
FT   VARSPLIC    390    421       NFS (IN ISOFORM NRG2-BETA).
FT   VARSPLIC    414    421       DRCQQFAMVNFSK (IN ISOFORM NTAK-BETA) -> VGYTG
FT   VARSPLIC    414    439       MISSING (IN ISOFORM NTAK-ALPHA2B).
FT   VARSPLIC                     ISOFORM NTAK-ALPHA2B)
FT                                HLGFELKEAEELYQKRVLITIGICVA -> SVLMDTPGTGV
```

```
FT   VARSPLIC    440    868       SSSQWSTSPSTLDLN (IN ISOFORM NRG2-ALPHA).
FT   CONFLICT    117    117       MISSING (IN ISOFORM NRG2-ALPHA).
FT   CONFLICT    117    117       S -> F (IN REF. 2).
FT   CONFLICT    724    724       R -> H (IN REF. 2).
SQ   SEQUENCE    868 AA;  93376 MW;  3C7D4D94DBE64DE2 CRC64;
```

```
Qy       1 HFKPCRDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRC 43
Db     357 HARCNEFAPKSYCVNGGVCYYIEGI--NQLSCKCPNGFFGQRC 397
```

```
Qy       1 HFKPCRDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRC 43
Db     357 HARCNEFAPKSYCVNGGVCYYIEGI--NQLSCKCPNGFFGQRC 397
```

```
Query Match             32.5%;  Score 90;  DB 1;  Length 868;
Best Local Similarity   34.9%;  Pred. No. 0.0019;
Matches     15;  Conservative    9;  Mismatches   17;  Indels    2;  Gaps    1;
```

```
RESULT 12
EREG_HUMAN
ID   EREG_HUMAN       STANDARD;       PRT;   169 AA.
AC   O14944;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   EPIREGULIN PRECURSOR.
GN   EREG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. AND CHARACTERIZATION.
RC   TISSUE=Colorectal adenocarcinoma;
RX   MEDLINE=97479200; PubMed=9337852;
RA   Toyoda H., Komurasaki T., Uchida D., Morimoto S.;
RT   "Distribution of mRNA for human epiregulin, a differentially expressed
RT   member of the epidermal growth factor family.";
RL   Biochem. J. 326:69-75(1997).
CC   -!- FUNCTION: MAY BE A MEDIATOR OF LOCALIZED CELL PROLIFERATION. AS A
CC       MITOGEN IT MAY STIMULATE CELL PROLIFERATION AND/OR ANGIOGENESIS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PRECURSOR FORM);
CC       EXTRACELLULAR (MATURE FORM).
CC   -!- TISSUE SPECIFICITY: IN NORMAL ADULTS EXPRESSED PREDOMINANTLY IN THE
CC       PLACENTA AND PERIPHERAL BLOOD LEUCOCYTES. HIGH LEVELS WERE DETECTED
CC       IN CARCINOMAS OF THE BLADDER, LUNG, KIDNEY AND COLON.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D30783; BAA22146.1; -.
DR   MIM; 602061;
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001336; EGF_1.
DR   PRINTS; PR000009; EGFTGF.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   Growth factor; Mitogen; Glycoprotein; EGF-like domain; Transmembrane;
KW   Signal.
FT   SIGNAL        1     29
FT   PROPEP       30     59       POTENTIAL.
FT   CHAIN        60    108       EPIREGULIN.
FT   PROPEP      109    169       REMOVED IN MATURE FORM (BY SIMILARITY).
FT   DOMAIN       60    119       POTENTIAL.
FT   TRANSMEM    120    140       POTENTIAL.
FT   DOMAIN      141    169       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      145    152       ARG/LYS-RICH (BASIC).
FT   CARBOHYD     47     47       N-LINKED (GLCNAC...) (POTENTIAL).
FT   DOMAIN       64    104       EGF-LIKE.
```

```
FT   DISULFID     68     81       BY SIMILARITY.
FT   DISULFID     76     92       BY SIMILARITY.
FT   DISULFID     94    103       BY SIMILARITY.
SQ   SEQUENCE   169 AA;  19044 MW;  17F3926ADFB2BDEE CRC64;

Qy        Query Match        32.3%;  Score 89.5;  DB 1;  Length 169;
          Best Local Similarity  40.5%;  Pred. No. 0.0049;
Matches 17;  Conservative 7;  Mismatches 15;  Indels 3;  Gaps 2;

Qy    5  CRDKDLAYCLNDEGCFVIETLTGSHKHCRCKEGYQGVRCDQF  46
          ||| :|:|  |||||||||.||:||||:||:||||:|||:||:|
Db   68  CSSDMNGYCLH-GQCIYLVDM--SQNYCRCEVGYTGVRCEHF  106
```

```
RESULT 13
GRFA_SFVKA
ID   GRFA_SFVKA     STANDARD;      PRT;     80 AA.
AC   P08441;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   GROWTH FACTOR.
OS   Shope fibroma virus (strain Kasza) (SFV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Leporipoxvirus.
OX   NCBI_TaxID=10272;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87172751; PubMed=3031480;
RA   Chang W., Upton C., Hu S.-L., Purchio A.F., McFadden G.;
RT   "The genome of Shope fibroma virus, a tumorigenic poxvirus, contains
RT   a growth factor gene with sequence similarity to those encoding
RT   epidermal growth factor and transforming growth factor alpha.";
RL   Mol. Cell. Biol. 7:535-540(1987).
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; M15921; AAA66873.1; -.
DR   PIR; A26723; EGVZSF.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001336; EGF_1.
DR   PRINTS; PR00009; EGFTGF.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
KW   EGF-like domain; Growth factor; Glycoprotein.
FT   DOMAIN       29     73       EGF-LIKE.
FT   DISULFID     33     47       BY SIMILARITY.
FT   DISULFID     41     61       BY SIMILARITY.
FT   DISULFID     63     72       BY SIMILARITY.
FT   CARBOHYD     44     44       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   80 AA;  9210 MW;  C48D30E878D2ED58 CRC64;

Qy        Query Match        30.7%;  Score 85;  DB 1;  Length 80;
          Best Local Similarity  39.2%;  Pred. No. 0.0085;
Matches 20;  Conservative 6;  Mismatches 17;  Indels 8;  Gaps 3;

Qy    1  HFKPCRDKDLAYCLNDEGCFVIE----TLTGSHKHCRCKEGYQGVRCDQFL  47
          | :||.||||||||:|| |       |:|| ||::||:||:|||||:|||
Db   29  HVKVCNHDYENYCLNNGTCFTIALDNYSIT---PFCVCRINYEGSRC-QFI  75
```

```
RESULT 14
NTC1_MOUSE
ID   NTC1_MOUSE     STANDARD;      PRT;   2531 AA.
AC   Q01705;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).
GN   NOTCH1 OR MOTCH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=93194170; PubMed=8449489;
RA   Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
RA   Copeland N.G., Gridley T.;
RT   "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT   homolog of Drosophila Notch.";
RL   Genomics 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 1551-2170 FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=93048835; PubMed=1425352;
RA   Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RA   Greenspan R.J., McMahon A.P., Gridley T.;
RT   "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT   suggests an important role in early postimplantation mouse
RT   development.";
RL   Development 115:737-744(1992).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; Z11886; CAA77941.1; -.
DR   HSSP; P00740; 1IXA.
DR   MGD; MGI:97363; Notch1.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF-2.
DR   InterPro; IPR000742; EGF_CA.
DR   InterPro; IPR001881; EGF_CA.
DR   InterPro; IPR001438; EGF-II.
DR   InterPro; IPR008800; Notch.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 35.
DR   Pfam; PF00066; notch; 3.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00248; ANK; 3.
DR   SMART; SM00179; EGF_CA; 23.
DR   SMART; SM00004; ANK_REPEAT; 2.
DR   SMART; SM00001; NL; 2.
DR   SMART; SM00181; EGF; 13.
DR   PROSITE; PS50088; ANK_REPEAT; 2.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS00187; EGF_CA; 21.
DR   PROSITE; PS01187; EGF_CA; ...
KW   Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Signal; Glycoprotein.
FT   SIGNAL        1     18       POTENTIAL.
```

```
FT   CHAIN        19   2531  NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN       19   1725  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     1726 1746  POTENTIAL.
FT   DOMAIN       1747 2531  CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       20   58    EGF-LIKE 1.
FT   DOMAIN       59   99    EGF-LIKE 2.
FT   DOMAIN       102  139   EGF-LIKE 3.
FT   DOMAIN       140  176   EGF-LIKE 4.
FT   DOMAIN       178  216   EGF-LIKE 5.
FT   DOMAIN       218  255   EGF-LIKE 6.
FT   DOMAIN       257  293   EGF-LIKE 7. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       295  333   EGF-LIKE 8. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       335  371   EGF-LIKE 9. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       372  410   EGF-LIKE 10. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       412  450   EGF-LIKE 11. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       452  488   EGF-LIKE 12. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       490  526   EGF-LIKE 13. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       528  564   EGF-LIKE 14. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       566  601   EGF-LIKE 15. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       603  639   EGF-LIKE 16. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       641  676   EGF-LIKE 17. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       678  714   EGF-LIKE 18. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       716  751   EGF-LIKE 19. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       753  789   EGF-LIKE 20. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       791  827   EGF-LIKE 21. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       829  867   EGF-LIKE 22. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       869  905   EGF-LIKE 23. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       907  943   EGF-LIKE 24. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       945  981   EGF-LIKE 25. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       983  1019  EGF-LIKE 26.
FT   DOMAIN       1021 1057  EGF-LIKE 27. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       1059 1095  EGF-LIKE 28.
FT   DOMAIN       1097 1143  EGF-LIKE 29.
FT   DOMAIN       1145 1181  EGF-LIKE 30. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       1183 1219  EGF-LIKE 31. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       1221 1265  EGF-LIKE 32. CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       1267 1305  EGF-LIKE 33.
FT   DOMAIN       1307 1346  EGF-LIKE 34.
FT   DOMAIN       1348 1384  EGF-LIKE 35.
FT   DOMAIN       1387 1426  EGF-LIKE 36.
FT   DOMAIN       1449 1462  CYS-RICH.
FT   REPEAT       1445 1480  LIN/NOTCH 1.
FT   REPEAT       1481 1522  LIN/NOTCH 2.
FT   REPEAT       1523 1562  LIN/NOTCH 3.
FT   REPEAT       1917 1947  ANK 1.
FT   REPEAT       1949 1979  ANK 2.
FT   REPEAT       1983 2012  ANK 3.
FT   REPEAT       2016 2045  ANK 4.
FT   REPEAT       2049 2078  ANK 5.
FT   DISULFID     24   37    BY SIMILARITY.
FT   DISULFID     31   46    BY SIMILARITY.
FT   DISULFID     63   74    BY SIMILARITY.
FT   DISULFID     68   87    BY SIMILARITY.
FT   DISULFID     89   98    BY SIMILARITY.
FT   DISULFID     106  117   BY SIMILARITY.
FT   DISULFID     111  127   BY SIMILARITY.
FT   DISULFID     129  138   BY SIMILARITY.
FT   DISULFID     144  155   BY SIMILARITY.
FT   DISULFID     149  164   BY SIMILARITY.
FT   DISULFID     166  175   BY SIMILARITY.
FT   DISULFID     182  195   BY SIMILARITY.
FT   DISULFID     189  204   BY SIMILARITY.
FT   DISULFID     206  215   BY SIMILARITY.
FT   DISULFID     222  233   BY SIMILARITY.
FT   DISULFID     227  243   BY SIMILARITY.
FT   DISULFID     245  254   BY SIMILARITY.
FT   DISULFID     261  272   BY SIMILARITY.
FT   DISULFID     266  281   BY SIMILARITY.
FT   DISULFID     283  292   BY SIMILARITY.
FT   DISULFID     299  312   BY SIMILARITY.
FT   DISULFID     306  321   BY SIMILARITY.
FT   DISULFID     323  332   BY SIMILARITY.
FT   DISULFID     339  350   BY SIMILARITY.
FT   DISULFID     344  359   BY SIMILARITY.
FT   DISULFID     361  370   BY SIMILARITY.
FT   DISULFID     376  387   BY SIMILARITY.
FT   DISULFID     381  398   BY SIMILARITY.
FT   DISULFID     400  409   BY SIMILARITY.
FT   DISULFID     416  429   BY SIMILARITY.
FT   DISULFID     423  438   BY SIMILARITY.
FT   DISULFID     440  449   BY SIMILARITY.
FT   DISULFID     456  467   BY SIMILARITY.
FT   DISULFID     461  476   BY SIMILARITY.
FT   DISULFID     478  487   BY SIMILARITY.
FT   DISULFID     494  505   BY SIMILARITY.
FT   DISULFID     499  514   BY SIMILARITY.
FT   DISULFID     516  525   BY SIMILARITY.
FT   DISULFID     532  543   BY SIMILARITY.
FT   DISULFID     537  552   BY SIMILARITY.
FT   DISULFID     554  563   BY SIMILARITY.
FT   DISULFID     570  580   BY SIMILARITY.
FT   DISULFID     575  589   BY SIMILARITY.
FT   DISULFID     591  600   BY SIMILARITY.
FT   DISULFID     607  618   BY SIMILARITY.
FT   DISULFID     612  627   BY SIMILARITY.
FT   DISULFID     629  638   BY SIMILARITY.
FT   DISULFID     645  655   BY SIMILARITY.
FT   DISULFID     650  664   BY SIMILARITY.
FT   DISULFID     666  675   BY SIMILARITY.
FT   DISULFID     682  693   BY SIMILARITY.
FT   DISULFID     687  702   BY SIMILARITY.
FT   DISULFID     704  713   BY SIMILARITY.
FT   DISULFID     720  730   BY SIMILARITY.
FT   DISULFID     725  739   BY SIMILARITY.
FT   DISULFID     741  750   BY SIMILARITY.
FT   DISULFID     757  768   BY SIMILARITY.
FT   DISULFID     762  777   BY SIMILARITY.
FT   DISULFID     779  788   BY SIMILARITY.
FT   DISULFID     795  806   BY SIMILARITY.
FT   DISULFID     800  815   BY SIMILARITY.
FT   DISULFID     817  826   BY SIMILARITY.
FT   DISULFID     833  844   BY SIMILARITY.
FT   DISULFID     838  855   BY SIMILARITY.
FT   DISULFID     857  866   BY SIMILARITY.
FT   DISULFID     873  884   BY SIMILARITY.
FT   DISULFID     878  893   BY SIMILARITY.
FT   DISULFID     895  904   BY SIMILARITY.
FT   DISULFID     911  922   BY SIMILARITY.
FT   DISULFID     916  931   BY SIMILARITY.
FT   DISULFID     933  942   BY SIMILARITY.
FT   DISULFID     987  998   BY SIMILARITY.
FT   DISULFID     992  1007  BY SIMILARITY.
FT   DISULFID     1009 1018  BY SIMILARITY.
FT   DISULFID     1025 1036  BY SIMILARITY.
FT   DISULFID     1030 1045  BY SIMILARITY.
FT   DISULFID     1047 1056  BY SIMILARITY.
FT   DISULFID     1063 1074  BY SIMILARITY.
FT   DISULFID     1068 1083  BY SIMILARITY.
```

```
RESULT 15
BTC_MOUSE
ID   BTC_MOUSE      STANDARD;      PRT;   177 AA.
AC   Q05928;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   BETACELLULIN PRECURSOR (BTC).

Query Match             30.0%;  Score 83;  DB 1;  Length 2531;
Best Local Similarity   41.7%;  Pred. No. 0.035;
Matches   15;  Conservative    4;  Mismatches    4;  Indels   13;  Gaps    1;

Qy        9  DLAYCLNDGECFVIETLGTGSHKHCRCKEGYQGVRCD  44
             | |  | ||| |  |:| | |:||  |  ||| || ||
Db     1064  DSAPCKNGGRCW----QTNTQYHCECRSGWTGVNCD  1095
```

```
GN   BTC OR BCN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 32-54; 64-71 AND 75-111.
RC   TISSUE=Pancreas;
RX   MEDLINE=93206093; PubMed=8456283;
RA   Shing Y., Christofori G., Ono Y., Sasada R.,
RA   Igarashi K., Folkman J.;
RT   "Betacellulin: a mitogen from pancreatic beta cell tumors.";
RL   Science 259:1604-1607(1993).
CC   -!- FUNCTION: POTENT MITOGEN FOR RETINAL PIGMENT EPITHELIAL CELLS
CC       AND VASCULAR SMOOTH MUSCLE CELLS. THE EFFECTS OF BETACELLULIN
CC       ARE PROBABLY MEDIATED BY THE EGF RECEPTOR AND OTHER RELATED
CC       RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PRECURSOR FORM);
CC       EXTRACELLULAR (MATURE FORM).
CC   -!- TISSUE SPECIFICITY: FOUND IN SEVERAL MOUSE TISSUES INCLUDING
CC       KIDNEY, UTERUS AND LIVER AS WELL AS IN BETA TUMOR CELL LINE AND
CC       MCF-7 CELLS. IT IS NOT DETECTED IN THE BRAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L08394; AAA05111.1; -.
DR   PIR; A37408; A37408.
DR   HSSP; P01135; 1YUF.
DR   MGD; MGI:99439; Btc.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001336; EGF_1.
DR   Pfam; PF00008; EGF; 1.
DR   PRINTS; PR00009; EGFTGF.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00001; EGF_like; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   Growth factor; Mitogen; Glycoprotein; EGF-like domain; Transmembrane;
KW   Signal.
FT   SIGNAL        1     31
FT   CHAIN        32    111       BETACELLULIN.
FT   PROPEP      112    177       REMOVED IN MATURE FORM.
FT   DOMAIN       32    118       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    119    139       POTENTIAL.
FT   DOMAIN      140    177       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       65    105       EGF-LIKE.
FT   DOMAIN      146    153       ARG/LYS-RICH (BASIC).
FT   DISULFID     69     82       BY SIMILARITY.
FT   DISULFID     77     93       BY SIMILARITY.
FT   DISULFID     95    104       BY SIMILARITY.
FT   CARBOHYD     34     34       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     42     42       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     52     52       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   177 AA;  19664 MW;  066B34F0E13F82B CRC64;
```

```
Query Match          29.8%;   Score 82.5;   DB 1;   Length 177;
Best Local Similarity 39.1%;   Pred. No. 0.0035;
Matches   18;  Conservative   9;  Mismatches   14;  Indels   5;  Gaps   3;

Qy      1  HFPKPCRDKDLAYCLNDGEC-FVIETLTGSHKHCRCKEGYQGVRCDQ  45
           ||  | |   | | |||  | ||    |   ||| | |  |  | | |
Db     65  HFSRCPKQYKHYCIH-GRCRFVVDEQTPS---CICEKGYFGARCER  106
```

Om protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: February 7, 2002, 12:59:52 ; Search time 22.8 Seconds
                 (without alignments)
                 301.526 Million cell updates/sec

Title: US-09-480-977-4
Perfect score: 277
Sequence: 1 HFKPCRDKDLAYCLNDGECF.........SHKHCRCKEGYQGVRCDQFL 47

Scoring table:
                 Gapop 10.0 , Gapext 0.5
                 BLOSUM62

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

                 Minimum DB seq length: 0
                 Maximum DB seq length: 2000000000

Database :
                 SPTREMBL_17:*
                 1: sp_archea:*
                 2: sp_bacteria:*
                 3: sp_fungi:*
                 4: sp_human:*
                 5: sp_invertebrate:*
                 6: sp_mammal:*
                 7: sp_mhc:*
                 8: sp_organelle:*
                 9: sp_phage:*
                 10: sp_plant:*
                 11: sp_rodent:*
                 12: sp_virus:*
                 13: sp_vertebrate:*
                 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	113.5	41.0	461	11	O35947	O35947 mesocricetu
2	111.5	40.3	298	11	Q9ESA9	Q9ESA9 rattus norv
3	111.5	40.3	695	11	Q9ESB0	Q9ESB0 rattus norv
4	104.5	37.7	241	6	Q07112	Q07112 bos taurus
5	102.5	37.0	111	11	Q9ESA8	Q9ESA8 rattus norv
6	102.5	37.0	136	11	Q9ESA7	Q9ESA7 rattus norv
7	102.5	37.0	256	11	Q9ESA6	Q9ESA6 rattus norv
8	102.5	37.0	317	11	Q9ESA3	Q9ESA3 rattus norv
9	102.5	37.0	323	11	Q9ESA2	Q9ESA2 rattus norv
10	102.5	37.0	342	11	Q9ESA1	Q9ESA1 rattus norv
11	102.5	37.0	700	11	Q9ESB1	Q9ESB1 rattus norv
12	102.5	37.0	782	11	Q9ESA5	Q9ESA5 rattus norv
13	92.5	33.4	2180	5	O01768	O01768 caenorhabdi
14	89	32.1	162	11	Q9Z0L5	Q9Z0L5 rattus norv
15	89	32.1	1241	11	Q9UKK5	Q9UKK5 homo sapien
16	89	32.1	1241	4	Q90KD4	Q90KD4 homo sapien
17	89	32.1	1241	4	Q9BZV3	Q9BZV3 homo sapien
18	88	31.8	1239	11	P70628	P70628 rattus norv
19	87	31.4	217	5	Q9VVJ6	Q9VVJ6 drosophila
20	87	31.4	597	11	O35727	O35727 mus musculu
21	85.5	30.9	1193	13	Q90819	Q90819 gallus gall
22	85.5	30.9	1218	4	O15122	O15122 homo sapien
23	85.5	30.9	1218	4	O15816	O15816 homo sapien
24	85.5	30.9	1218	4	O14902	O14902 homo sapien
25	85.5	30.9	1227	4	P78504	P78504 homo sapien
26	84.5	30.5	177	11	Q9JJM4	Q9JJM4 rattus norv
27	84.5	30.5	1218	11	Q9QXX0	Q9QXX0 mus musculu
28	84.5	30.5	1219	11	O63722	O63722 rattus norv
29	83.5	30.1	127	12	Q90F3	Q90F3 ectromelia
30	83	30.0	162	11	O61521	O61521 mus musculu
31	83	30.0	861	11	Q9WU58	Q9WU58 sp. mot
32	83	30.0	862	11	Q99JC2	Q99JC2 mus musculu
33	82.5	29.8	445	5	Q9W3W5	Q9W3W5 drosophila
34	82	29.6	264	5	O20559	O20559 caenorhabdi
35	80.5	29.1	125	12	O41504	O41504 cowpox viru
36	80.5	29.1	125	12	Q9J524	Q9J524 fowlpox vir
37	80.5	29.1	260	4	O95898	O95898 homo sapien
38	80.5	29.1	2653	5	O25253	O25253 lucilia cup
39	80	28.9	907	5	Q9XTS9	Q9XTS9 caenorhabdi
40	79.5	28.7	138	12	O41506	O41506 bean 58058
41	79.5	28.7	138	12	P87605	P87605 cowpox viru
42	79.5	28.7	140	12	O57166	O57166 vaccinia vi
43	79.5	28.7	140	12	O86607	O86607 vaccinia vi
44	79.5	28.7	140	12	Q9JFH4	Q9JFH4 vaccinia vi
45	79.5	28.7	1581	13	O73809	O73809 fugu rubrip

ALIGNMENTS

RESULT   1
O35947

O35947   PRELIMINARY;   PRT;   461 AA.
ID   O35947
AC   O35947;
DT   01-JAN-1998 (TREMBLrel. 05, Created)
DT   01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE   PRO-NEUREGULIN-1, ISOFORM ALPHA 2B PRECURSOR.
GN   NRG1 OR NDF.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA2B), AND SEQUENCE OF 64-81.
RC   TISSUE=EMBRYO;
RX   MEDLINE=98196996; PubMed=9537646;
RA   Velasco J.A., Feijoo E., Avila M.A., Notario V.;
RT   "Secretion of neuregulin factor-like polypeptides by cph-
RT   transformed fibroblasts: cloning and characterization of Syrian
RT   hamster neuregulin cDNAs.";
RL   Mol. Carcinog. 21:156-163(1998).
CC   -!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE
CC   RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
CC   RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC   ACTIVATION OF THE ERBB RECEPTORS. MAY PLAY AN IMPORTANT ROLE IN
CC   PROVIDING GROWTH ADVANTAGE IN NEOPLASTIC CELLS.
CC   -!- SUBUNIT: THE CYTOPLASMIC DOMAIN INTERACTS WITH THE LIM DOMAIN
CC   REGION OF LIMK1 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: EXISTS AS TYPE I MEMBRANE PROTEIN AND AS A
CC   PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE MEMBRANE-
CC   BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS ARE PRODUCED BY
CC   ALTERNATIVE SPLICING. THE SEQUENCE SHOWN IS THAT OF ISOFORM
CC   ALPHA2B/CLONE PM3.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHER LEVEL AFTER NEOPLASTIC
CC   TRANSFORMATION OF CELLS.
CC   -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC   OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC   PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC   DIMERIZATION (BY SIMILARITY).

CC -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC   DOMAIN.
CC -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
CC   EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC   FORM (BY SIMILARITY).
CC -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES PROTEOLYTIC CLEAVAGE (BY
CC   SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
DR EMBL; U96612; AAB71812.1; -.
DR HSSP; Q12784; 1HRE.
DR Pfam; PF00008; EGF; 1.
DR InterPro; IPR002154; Neuregulin.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR000561; EGF-like.
DR PRINTS; PR01089; NEUREGULIN.
DR Pfam; PF02158; Neuregulin; 1.
DR Pfam; PF00047; Ig; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR SMART; SM00408; IG2; 1.
DR SMART; SM00181; EGF; 1.
KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW Transmembrane; Alternative splicing.
FT PROPEP   1   13   BY SIMILARITY.
FT CHAIN   14   461   NEUREGULIN-1.
FT CHAIN   14   241   PRO-NEUREGULIN-1, MEMBRANE-BOUND FORM.
FT DOMAIN   14   242   EXTRACELLULAR (POTENTIAL).
FT TRANSMEM   243   265   INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT DOMAIN   266   461   CYTOPLASMIC (POTENTIAL).
FT DOMAIN   50   119   IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN   165   177   SER/THR-RICH.
FT DOMAIN   178   222   EGF-LIKE.
FT DISULFID   57   112   BY SIMILARITY.
FT DISULFID   112   196   BY SIMILARITY.
FT DISULFID   182   196   BY SIMILARITY.
FT DISULFID   190   210   BY SIMILARITY.
FT DISULFID   212   221   BY SIMILARITY.
FT CARBOHYD   73   73   N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD   120   120   N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD   126   126   N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD   164   164   N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE   461 AA; 50890 MW; 93C59560F7148336 CRC64;

Qy   1 HFKPCRDKDLAYCLNDGECFVIELTGSHKH-CRCKEGYQGVRCDQ 45
                 |:|: ||| ||||| |:|::||:||:|:|:|||
Db   178 HLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCKCQPGFTGARCTE 223

Query Match   41.0%; Score 113.5; DB 11; Length 461;
Best Local Similarity   34.8%;
Matches 16; Conservative 14; Mismatches 15; Indels   1; Gaps   1;

RESULT 2
ID Q9ESA9   PRELIMINARY;   PRT;   298 AA.
AC Q9ESA9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SMDF NEUREGULIN ALPHA 2B (FRAGMENT).
GN NRG1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BDIX;
RA Carroll S.L., Anderson K.D., Frohert P.W.;
RT "Structural and Functional Diversity of SMDF Neuregulin Splice

RESULT 3
ID Q9ESB0   PRELIMINARY;   PRT;   695 AA.
AC Q9ESB0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SMDF NEUREGULIN ALPHA 2A.
GN NRG1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BDIX;
RA Carroll S.L., Anderson K.D., Frohert P.W.;
RT "Structural and Functional Diversity of SMDF Neuregulin Splice
RT Variants Expressed in the Adult Rat Nervous System.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF194439; AAG28428.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002154; Neuregulin.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF02158; Neuregulin; 1.
DR PRINTS; PR01089; NEUREGULIN.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
SQ SEQUENCE   695 AA; 75646 MW; 5277F2CBA2FB6878 CRC64;

Qy   1 HFKPCRDKDLAYCLNDGECFVIELTGSHKH-CRCKEGYQGVRCDQ 45
                 |:|: ||| ||||| |:|::||:||:|:|:|||
Db   234 HLIKCAEKEKTFCVNGGECFTVKDLSNPSRYLCKCQPGFTGARCTE 279

Query Match   40.3%; Score 111.5; DB 11; Length 695;
Best Local Similarity   34.8%;
Matches 16; Conservative 13; Mismatches 16; Indels   1; Gaps   1;

RESULT 4
ID Q07112   PRELIMINARY;   PRT;   241 AA.
AC Q07112;

RESULT 3
ID Q9ESB0   PRELIMINARY;   PRT;   695 AA.
AC Q9ESB0;
DR EMBL; AF194440; AAG28429.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002154; Neuregulin.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF02158; Neuregulin; 1.
DR PRINTS; PR01089; NEUREGULIN.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
SQ SEQUENCE   298 AA; 32851 MW; BD76F014C2B33026 CRC64;

Qy   1 HFKPCRDKDLAYCLNDGECFVIELTGSHKH-CRCKEGYQGVRCDQ 45
                 |:|: ||| ||||| |:|::||:||:|:|:|||
Db   48 HLIKCAEKEKTFCVNGGECFTVKDLSNPSRYLCKCQPGFTGARCTE 93

Query Match   40.3%; Score 111.5; DB 11; Length 298;
Best Local Similarity   34.8%;
Matches 16; Conservative 13; Mismatches 16; Indels   1; Gaps   1;

```
AC   Q07112;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   NEUREGULIN-1, GLIAL GROWTH FACTOR 5 ISOFORM PRECURSOR (GGFBPP5).
GN   NRG1 OR GGF.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=POSTERIOR PITUITARY;
RX   MEDLINE=93205115; PubMed=8096067;
RA   Marchionni M.A., Goodearl A.D.J., Chen M.S., Bermingham-McDonogh O.,
RA   Kirk C., Hendricks M., Danehy F., Misumi D., Sudhalter J.,
RA   Kobayashi K., Wroblewski D., Lynch C., Baldassare M., Hiles I.,
RA   Davis J.B., Hsuan J.J., Totty N.F., Otsu M., McBurney R.N.,
RA   Waterfield M.D., Stroobant P., Gwynne D.;
RT   "Glial growth factors are alternatively spliced erbB2 ligands
RT   expressed in the nervous system.";
RL   Nature 362:312-318(1993).
CC   -!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE
CC       RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
CC       RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC       ACTIVATION OF THE ERBB RECEPTORS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED (PROBABLE).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN NERVOUS SYSTEM: SPINAL CORD AND
CC       BRAIN.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST FIVE DIFFERENT ISOFORMS; GGFBPP1,
CC       GGFBPP2, GGFBPP3, GGFBPP4 AND GGFBPP5 (SHOWN HERE); ARE PRODUCED
CC       BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
DR   EMBL; L12259; AAA30540.1; -.
DR   HSSP; Q12784; 1HRE.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00408; IG2; 1.
DR   SMART; SM00409; IG; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS01185; EGF_2; 1.
KW   Growth factor; EGF-like domain; Immunoglobulin domain;
KW   Alternative splicing.
FT   PROPEP        1     19
FT   CHAIN        20    241       NEUREGULIN-1, GLIAL GROWTH FACTOR 5
FT                                ISOFORM.
FT   DOMAIN       50    119       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      165    177       SER/THR-RICH.
FT   DOMAIN      178    222       EGF-LIKE DOMAIN.
FT   DISULFID    182    196       BY SIMILARITY.
FT   DISULFID    190    210       BY SIMILARITY.
FT   DISULFID    212    221       BY SIMILARITY.
FT   VARSPLIC    134    156       EITTGMPASTETAVVSSESPIRI -> GKRCLLRAISOSLR
FT                                GVIRVCGHT (IN ISOFORM GGFBPP1).
FT   VARSPLIC    157    241       MISSING (IN ISOFORM GGFBPP1).
SQ   SEQUENCE   241 AA;  25955 MW;  BF571297E8DA9796 CRC64;

Qy
Db

Query Match             37.7%;  Score 104.5;  DB 6;  Length 241;
Best Local Similarity   31.2%;  Pred. No. 2.8e-06;
Matches  15;  Conservative  14;  Mismatches  18;  Indels  1;  Gaps  1;

Qy    1  HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL  47
            |.|: .||.| :|||||||||||||| |||:| :|.| ||.|:|| ||
Db  178  HLVKCAEKEKTFCVNGECFMVKDLSNPSRYLCKCPNEFTGDRCQNYV  225
```

```
RESULT 5

ID   Q9ESA8      PRELIMINARY;      PRT;   111 AA.
AC   Q9ESA8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SMDF NEUREGULIN BETA 2 (FRAGMENT).
GN   NRG1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RC   TISSUE=AXOTOMIZED LUMBAR DORSAL ROOT GANGLION/SPINAL CORD;
RA   Carroll S.L., Anderson K.D., Frohnert P.W.;
RT   "Structural and Functional Diversity of SMDF Neuregulin Splice
RT   Variants Expressed in the Adult Rat Nervous System.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF194441; AAG28430.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000886; ER_target.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
FT   NON_TER       1      1
FT   NON_TER     111    111
SQ   SEQUENCE   111 AA;  12198 MW;  CC8BB8705B4C9F8C CRC64;

Qy
Db

Query Match             37.0%;  Score 102.5;  DB 11;  Length 111;
Best Local Similarity   31.2%;  Pred. No. 2.5e-06;
Matches  15;  Conservative  13;  Mismatches  19;  Indels  1;  Gaps  1;

Qy    1  HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL  47
Db   54  HLIKCAEKEKTFCVNGGECFTVKDLSNPSRYLCKCPNEFTGDRCQNYV  101
```

```
RESULT 6

ID   Q9ESA7      PRELIMINARY;      PRT;   136 AA.
AC   Q9ESA7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SMDF NEUREGULIN BETA 4 (FRAGMENT).
GN   NRG1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RC   TISSUE=AXOTOMIZED LUMBAR DORSAL ROOT GANGLION/SPINAL CORD;
RA   Carroll S.L., Anderson K.D., Frohnert P.W.;
RT   "Structural and Functional Diversity of SMDF Neuregulin Splice
RT   Variants Expressed in the Adult Rat Nervous System.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF194442; AAG28431.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
FT   NON_TER       1      1
FT   NON_TER     136    136
SQ   SEQUENCE   136 AA;  15235 MW;  116CBC91D6E8AEF5 CRC64;
```

RESULT 7

Query Match
Best Local Similarity 37.0%; Score 102.5; DB 11; Length 136;
         Conservative 13; Mismatches 19; Indels 1; Gaps 1;
Matches 15;

```
Qy    1 HFKPCRDKLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL 47
          |:|:| ||||| :|:|| ::|:| |:|| || .:
Db   54 HLIKCAEKEKTFCVNGGECFTVKDLSNPSRYLCKCPNEFTGDRCQNYV 101
```

ID   Q9ESA6   PRELIMINARY;   PRT;   256 AA.
AC   Q9ESA6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SMDF NEUREGULIN BETA 3 (FRAGMENT).
GN   NRG1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RC   TISSUE=AXOTOMIZED LUMBAR DORSAL ROOT GANGLION/SPINAL CORD;
RA   Carroll S.L., Anderson K.D., Frohnert P.W.;
RT   "Structural and Functional Diversity of SMDF Neuregulin Splice
RT   variants Expressed in the Adult Rat Nervous System.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF194443; AAG28432.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00001; EGF_like; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
FT   NON_TER   1
FT   NON_TER   1
SQ   SEQUENCE   256 AA; 27735 MW; C8C08DEA68D33E39 CRC64;

RESULT 8

Query Match
Best Local Similarity 37.0%; Score 102.5; DB 11; Length 256;
         Conservative 13; Mismatches 19; Indels 1; Gaps 1;
Matches 15;

```
Qy    1 HFKPCRDKLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL 47
          |:|:| ||||| :|:|| ::|:| |:|| || .:
Db  193 HLIKCAEKEKTFCVNGGECFTVKDLSNPSRYLCKCPNEFTGDRCQNYV 240
```

ID   Q9ESA3   PRELIMINARY;   PRT;   317 AA.
AC   Q9ESA3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   GLIAL GROWTH FACTOR GGF BETA 2 (FRAGMENT).
GN   NRG1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RC   TISSUE=AXOTOMIZED LUMBAR DORSAL ROOT GANGLION/SPINAL CORD;
RA   Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,
RA   Frohnert P.W.;
RT   "Structural and Functional Diversity of Glial Growth Factor Isoforms
RT   Expressed in Regenerating Peripheral Nerve and Associated Neurons.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC   DOMAIN.
DR   EMBL; AF194996; AAG28450.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; Ig; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
FT   NON_TER   1
SQ   SEQUENCE   323 AA; 35558 MW; C7DF153A939A80C8 CRC64;

RESULT 9

Query Match
Best Local Similarity 37.0%; Score 102.5; DB 11; Length 317;
         Conservative 13; Mismatches 19; Indels 1; Gaps 1;
Matches 15;

```
Qy    1 HFKPCRDKLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL 47
          |:|:| ||||| :|:|| ::|:| |:|| || .:
Db  260 HLIKCAEKEKTFCVNGGECFTVKDLSNPSRYLCKCPNEFTGDRCQNYV 307
```

ID   Q9ESA2   PRELIMINARY;   PRT;   323 AA.
AC   Q9ESA2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   GLIAL GROWTH FACTOR GGF BETA 3 (FRAGMENT).
GN   NRG1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RC   TISSUE=AXOTOMIZED LUMBAR DORSAL ROOT GANGLION/SPINAL CORD;
RA   Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,
RA   Frohnert P.W.;
RT   "Structural and Functional Diversity of Glial Growth Factor Isoforms
RT   Expressed in Regenerating Peripheral Nerve and Associated Neurons.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC   DOMAIN.
DR   EMBL; AF194995; AAG28449.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000886; ER_target.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
FT   NON_TER   1
FT   NON_TER   1
SQ   SEQUENCE   317 AA; 34785 MW; 4487FA3E9CD876B9 CRC64;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL 47
        | :|:| ||||:| |||| :: |:| ::| :| ::
Db 260 HLIKCAEKEKTFCVNGGECFTVKDLSNPSRYLCKCPNEFTGDRCQNYV 307

RESULT 10
ID   Q9ESA1   PRELIMINARY;   PRT;   342 AA.
AC   Q9ESA1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   GLIAL GROWTH FACTOR GGF BETA 4 (FRAGMENT).
GN   NRG1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RC   TISSUE=AXOTOMIZED LUMBAR DORSAL ROOT GANGLION/SPINAL CORD;
RA   Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,
RA   Frohnert P.W.;
RT   "Structural and Functional Diversity of Glial Growth Factor Isoforms
RT   Expressed in Regenerating Peripheral Nerve and Associated Neurons.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC   DOMAIN.
DR   EMBL; AF194997; AAG28451.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00047; Ig; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
FT   NON_TER   1   1
FT   NON_TER   342   342
SQ   SEQUENCE   342 AA; 37836 MW; 8BE36FC836553124 CRC64;

Query Match   37.0%;
Best Local Similarity   31.2%;
Matches   15;   Conservative   13;   Mismatches   19;   Indels   1;   Gaps   1;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL 47
        | :|:| ||||:| |||| :: |:| ::| :| ::
Db 260 HLIKCAEKEKTFCVNGGECFTVKDLSNPSRYLCKCPNEFTGDRCQNYV 307

RESULT 11
ID   Q9ESB1   PRELIMINARY;   PRT;   700 AA.
AC   Q9ESB1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SMDF NEUREGULIN BETA 1A.
GN   NRG1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RA   Carroll S.L., Anderson K.D., Frohnert P.W.;
RT   "Structural and Functional Diversity of SMDF Neuregulin Splice

---

RT   Variants Expressed in the Adult Rat Nervous System.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF194438; AAG28427.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   PRINTS; PR01089; NEUREGULIN.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00001; EGF_like; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
SQ   SEQUENCE   700 AA; 76386 MW; 2F8111B17ECC49DA CRC64;

Query Match   37.0%;   Score 102.5;   DB 11;   Length 700;
Best Local Similarity   31.2%;   Pred. No. 1.4e-05;
Matches   15;   Conservative   13;   Mismatches   19;   Indels   1;   Gaps   1;

Qy   1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL 47
        | :|:| ||||:| |||| :: |:| ::| :| ::
Db 234 HLIKCAEKEKTFCVNGGECFTVKDLSNPSRYLCKCPNEFTGDRCQNYV 281

RESULT 12
ID   Q9ESA5   PRELIMINARY;   PRT;   782 AA.
AC   Q9ESA5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   GLIAL GROWTH FACTOR BETA 1A (FRAGMENT).
GN   NRG1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=SPINAL CORD/BRAIN STEM;
RA   Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,
RA   Frohnert P.W.;
RT   "Structural and Functional Diversity of Glial Growth Factor Isoforms
RT   Expressed in Regenerating Peripheral Nerve and Associated Neurons.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC   DOMAIN.
DR   EMBL; AF194993; AAG28433.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR002154; Neuregulin.
DR   InterPro; IPR002114; PTS_HPr_Ser.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00047; Ig; 1.
DR   Pfam; PF02158; Neuregulin; 1.
DR   PRINTS; PR01089; NEUREGULIN.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
FT   NON_TER   1   1
SQ   SEQUENCE   782 AA; 86036 MW; F6174A68F4E27BDE CRC64;

Query Match   37.0%;   Score 102.5;   DB 11;   Length 782;
Best Local Similarity   31.2%;   Pred. No. 1.6e-05;
Matches   15;   Conservative   13;   Mismatches   19;   Indels   1;   Gaps   1;

```
Qy      1 HFKPCRDKDLAYCLNDGECFVIETLTGSHKH-CRCKEGYQGVRCDQFL 47
            : ||:|:|| |||| ||||:|:|| :|:|:||: : ||   || ::
Db    316 HLIKCAEKEKTPCVNGGECFTVKDLSNPSRYLCKCPNEFTGDRCQNYV 363
```

RESULT 13

ID   OO1768   PRELIMINARY;   PRT;   2180 AA.
AC   OO1768;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 241.7 KDA PROTEIN T21E3.3 IN CHROMOSOME I.
GN   T21E3.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Du Z., Le T.T.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
CC   (LDLRA) DOMAIN.
DR   EMBL; AF003133; AAB54138.1; -.
DR   HSSP; Q07954; 1CR8.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR000033; Ldl_receptor_rep.
DR   InterPro; IPR002172; LDL_recept_A.
DR   Pfam; PF00008; EGF; 9.
DR   Pfam; PF00057; ldl_recept_a; 18.
DR   Pfam; PF00058; ldl_recept_b; 8.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   SMART; SM00181; EGF; 10.
DR   SMART; SM00192; LDLa; 20.
DR   SMART; SM00135; LY; 7.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01209; LDLRA_1; 10.
DR   PROSITE; PS50068; LDLRA_2; 21.
KW   EGF-like domain; Glycoprotein; Hypothetical protein.
SQ   SEQUENCE 2180 AA; 241705 MW; C26419F456A60D13 CRC64;

Query Match   33.4%; Score 92.5; DB 5; Length 2180;
Best Local Similarity   42.9%; Pred. No. 0.0092;
Matches   18; Conservative   8; Mismatches   7; Indels   9; Gaps   3;

```
Qy      5 CRDKDLAYCLNDGECFVIETLT-GSHKHCRCKEGYQGVRCDQ 45
            | | :|| ||:|:|| |:|:| :|:|:||:|:|:|:|||:|
Db   1906 CDD----YCTNNSKC---TITNGTHFECDCKPGFKGLRCEQ 1939
```

RESULT 14

ID   Q9Z0L5   PRELIMINARY;   PRT;   162 AA.
AC   Q9Z0L5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EPIREGULIN PRECURSOR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 56-75, AND CHARACTERIZATION.
RC   TISSUE=AORTIC SMOOTH MUSCLE;
RX   MEDLINE=99145602; PubMed=9990076;
RA   Taylor D.S., Cheng X., Pawlowski J.E., Wallace A.R., Ferrer P.,

RA   Molloy C.J.;
RT   "Epiregulin is a potent vascular smooth muscle cell-derived mitogen
RT   induced by angiotensin II, endothelin-1, and thrombin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:1633-1638(1999).
CC   -!- FUNCTION: MAY BE A MEDIATOR OF LOCALIZED CELL PROLIFERATION. AS A
CC   MITOGEN IT MAY STIMULATE CELL PROLIFERATION AND/OR ANGIOGENESIS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PRECURSOR FORM);
CC   EXTRACELLULAR (MATURE FORM).
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
DR   EMBL; AF074952; AAD10631.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001336; EGF_1.
DR   PRINTS; PR00009; EGFPGF.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   Growth factor; Mitogen; Glycoprotein; EGF-like domain; Transmembrane;
KW   Signal.
FT   SIGNAL   1   22   POTENTIAL.
FT   PROPEP   23   55   POTENTIAL.
FT   CHAIN   56   101   EPIREGULIN.
FT   PROPEP   102   162   REMOVED IN MATURE FORM (BY SIMILARITY).
FT   DOMAIN   56   112   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM 113   133   POTENTIAL.
FT   DOMAIN   134   162   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN   138   145   ARG/LYS-RICH (BASIC).
FT   CARBOHYD   40   40   N-LINKED (GLCNAC...) (POTENTIAL).
FT   DOMAIN   57   97   EGF-LIKE.
FT   DISULFID   61   74   BY SIMILARITY.
FT   DISULFID   69   85   BY SIMILARITY.
FT   DISULFID   87   96   BY SIMILARITY.
SQ   SEQUENCE   162 AA;   18404 MW; EA660DE2B34990C4 CRC64;

Query Match   32.1%; Score 89; DB 11; Length 162;
Best Local Similarity   45.0%; Pred. No. 0.00023;
Matches   18; Conservative   7; Mismatches   11; Indels   4; Gaps   3;

```
Qy      7 DKDLAYCLNDGECFVIETLTGSHKHCRCKEGYQGVRCDQF 46
            | || ||:|:|| |:|:|  | |||:||:|:|:|:|| |
Db     64 DMD-GYCLH-GHCIYLVDM--SEKYCRCEVGYTGLRCEHF 99
```

RESULT 15

ID   Q9UKK5   PRELIMINARY;   PRT;   1241 AA.
AC   Q9UKK5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPACRCAN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Acharya S., Foletta V.C., Lee J.W., Rayborn M.E., Rodriguez I.R.,
RA   Young W.S., Hollyfield J.G.;
RT   "SPACRCAN, a novel human interphotoreceptor matrix hyaluronan-binding
RT   proteoglycan synthesized by photoreceptors and pinealocytes.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF157624; AAF13154.1; -.
DR   InterPro; IPR000082; SEA.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF01390; SEA; 2.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   SMART; SM00200; SEA; 2.
SQ   SEQUENCE 1241 AA; 138605 MW; 1F3AE63D839F8858 CRC64;

Query Match   32.1%; Score 89; DB 4; Length 1241;
Best Local Similarity   34.6%; Pred. No. 0.0016;

```
Matches   18;  Conservative  11;  Mismatches  13;  Indels  10;  Gaps  3;

Qy      3 RPCR---DKDLAYCLNDGECFVIETLTGSHKHCRCKEG----YQGVRCDQFL 47
          :||: |   :|||||:|   |||:|   |:|  |||||:|
Db   1048 RPCQSLCDLQPDFCLNDGKC---DIMPGHGAICRCRVGENWWYRGKHCEEFV 1096
```

Search completed: February 7, 2002, 13:03:00
Job time: 188 sec